Mon. Not. R. Astron. Soc. **297,** 692–712 (1998)

The bias field of dark matter haloes

Paolo Catelan,[1,2] Francesco Lucchin,[3] Sabino Matarrese[4] and Cristiano Porciani[5]

[1]*Theoretical Astrophysics Center, Juliane Maries Vej 30, DK-2100 Copenhagen Ø, Denmark*
[2]*Department of Physics, Astrophysics, Nuclear Physics Laboratory, Keble Road, Oxford OX1 3RH*
[3]*Dipartimento di Astronomia, Università di Padova, vicolo dell'Osservatorio 5, I-35122 Padova, Italy*
[4]*Dipartimento di Fisica* Galileo Galilei, *Università di Padova, via Marzolo 8, I-35131 Padova, Italy*
[5]*SISSA, Scuola Internazionale di Studi Superiori Avanzati, via Beirut 2-4, I-34014 Trieste, Italy*

Accepted 1998 January 26. Received 1998 January 26; in original form 1997 August 13

ABSTRACT

This paper presents a stochastic approach to the clustering evolution of dark matter haloes in the Universe. Haloes, identified by a Press–Schechter-type algorithm in Lagrangian space, are described in terms of 'counting fields', acting as non-linear operators on the underlying Gaussian density fluctuations. By ensemble-averaging these counting fields, the standard Press–Schechter mass function as well as analytic expressions for the halo correlation function and corresponding bias factors of linear theory are obtained, extending the recent results by Mo & White. The non-linear evolution of our halo population is then followed by solving the continuity equation, under the sole hypothesis that haloes move by the action of gravity. This leads to an exact and general formula for the *bias field* of dark matter haloes, defined as the local ratio between their number density contrast and the mass density fluctuation. Besides being a function of position and 'observation' redshift, this random field depends upon the mass and formation epoch of the objects and is both non-linear and non-local. The latter features are expected to leave a detectable imprint on the spatial clustering of galaxies, as described, for instance, by statistics like the bispectrum and the skewness. Our algorithm may have several interesting applications, among which is the possibility of generating mock halo catalogues from low-resolution *N*-body simulations.

Key words: galaxies: clusters: general – galaxies: evolution – galaxies: formation – galaxies: haloes – cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

The theory proposed by Press & Schechter (1974, hereafter PS) to obtain the relative abundance of matter condensations in the Universe has strongly influenced all later studies on the statistical properties of dark matter haloes and led to a large variety of extensions, improvements and applications. Actually, already in the sixties, Doroshkevich (1967) had derived the mass distribution function for 'newly generated cosmic objects', completely analogous to the PS one; he had also clearly pointed out the existence of what has been later referred to as the *cloud-in-cloud problem* (e.g. Bardeen et al. 1986). The 'Press–Schechter model', which is based on the gravitational instability hypothesis, is now considered as one of the cornerstones of the hierarchical scenario for structure formation in the Universe. It shows, in fact, how gravitational instability makes more and more massive condensations grow by the aggregation of smaller units, provided only that the initial density fluctuation field contains enough power on small scales. The main drawback of the original PS model is indeed the cloud-in-cloud problem, i.e., the fact that their procedure selects bound objects of given mass that can have been already included in larger mass condensations of the same catalogue. The problem was later solved by several authors (Peacock & Heavens 1990; Bond et al. 1991; Cole 1991) according to the so-called 'excursion set' approach, by calculating the distribution of first-passage 'times' across the collapse threshold for suitably defined random walks. Lacey & Cole (1993, 1994) implemented these ideas to study the merger rates of virialized haloes in hierarchical models of structure formation.

An important aspect of the PS model is that, being entirely based on linear theory, suitably extrapolated to the collapse time of spherical perturbations, it is, by definition, local in Lagrangian space. While this Lagrangian aspect of the theory does not have immediate implications for the study of the mass function of dark matter haloes, it is, instead, of crucial importance for their spatial clustering properties. This point was recognized by Cole & Kaiser (1989) and, more recently, by Mo & White (1996, hereafter MW), who proposed a bias model for halo clustering in Eulerian space, by a suitable extension of the original PS algorithm for the mass function. With their formalism MW studied the clustering of dark matter haloes with different formation epochs (see also Mo, Jing & White 1996). The comparison of their theoretical predictions with the spatial distribution of haloes obtained by a *friends-of-friends* group finder and a spherical overdensity criterion in numerical simulations proved extremely successful.

These very facts imply that there exists a local version of the PS algorithm, providing a *mapping* betwen points of Lagrangian space and the haloes in embryo which will come into existence at the various epochs. For a given realization of the initial density field, the PS mapping is such that, at a fixed redshift z, each Lagrangian point $\boldsymbol{q}$ can be assigned to a matter clump of some mass M, identified by a suitable Lagrangian filter, which is collapsing at the epoch $z_{\rm f}=z$. One can therefore exploit the existence of this mapping to assign a stochastic halo process, our *halo counting field* below, to each point $\boldsymbol{q}$. This will be the starting point of our analysis.

What the PS Ansatz cannot account for is the fact that the fluid elements are moved apart by gravity, so that the halo which the PS mapping assigns to the fluid patch with Lagrangian coordinate $\boldsymbol{q}$ is not going to collapse in the same position, i.e., at $\boldsymbol{x}=\boldsymbol{q}$, but, rather in the Eulerian point $\boldsymbol{x}(\boldsymbol{q}, z)=\boldsymbol{q}+\boldsymbol{S}(\boldsymbol{q}, z)$, with $\boldsymbol{S}(\boldsymbol{q}, z)$ the displacement vector, corresponding to the Lagrangian one at the epoch $z=z_{\rm f}(\boldsymbol{q}, M)$ of its collapse. This fact, while not affecting in any way the PS result for the mean mass function, as the average halo abundance cannot change by scrambling the objects, sensibly modifies their spatial clustering properties. Modelling the latter effect is one of the main purposes of the present work. In their derivation of the Eulerian halo bias MW took into some account this problem by allowing for the local compression, or expansion, of the volumes where the haloes are located, an effect which is of crucial importance for the derivation of the correct halo density contrast. Their derivation, however, is formally flawed by the fact that they only deal with mean halo number densities, so that they are forced to define the bias in terms of them. For reasons to be shown below, however, this heuristic treatment can be put on sounder statistical grounds, by applying a suitable coarse-graining procedure.

Of course, the PS model has its own limitations. The comparison of its predictions for the mass function with the outputs of N-body simulations (e.g. Efstathiou et al. 1988; Gelb & Bertschinger 1994; Lacey & Cole 1994), while surprisingly successful in its general trends, given the simplicity of the assumptions, showed a number of problems. Gelb & Bertschinger, for instance, found that the simulated haloes are generally less massive than predicted, the reason being that merging does not erase substructure in large haloes as fast as required by the PS recipe.

There have been many attempts to improve the original PS model. If cosmic structures preferentially formed at the peaks of the initial density fluctuation field, this would affect their mean mass function (Bardeen et al. 1986; Bond 1988; Colafrancesco, Lucchin & Matarrese 1989; Peacock & Heavens 1990; Manrique & Salvador-Solé 1995, 1996). Bond & Myers (1996) developed a *peak-patch* picture of cosmic structure formation, according to which virialized objects are identified with suitable peaks of the Lagrangian density field. The peak-patch collapse dynamics is then followed in terms of the homogeneous ellipsoid model, which allows for the influence of the external tidal field, while the Zel'dovich approximation (Zel'dovich 1970) is used for the external peak-patch dynamics. The effects of non-spherical collapse on the shape of the mass distribution were studied by Monaco (1995). Lee & Shandarin (1997) analytically derived the mass function of gravitationally bound objects in the frame of the Zel'dovich approximation.

We prefer here to follow the simple lines of the PS theory to set up the 'initial conditions' for our stochastic approach to the evolution of halo clustering. Nevertheless, one should keep in mind that our approach is flexible enough to accept many levels of improvement in the treatment of the Lagrangian initial conditions.

A relevant part of the following analysis will be devoted to the study of the evolution of halo clustering away from the linear regime. It turns out that the problem can be solved exactly in terms of the evolved mass density. An important result of this analysis is that the general Eulerian bias factor, defined as the local ratio between the halo density contrast and the mass fluctuation field, is both non-linear and non-local. The latter property follows directly from our selection criterion of candidate haloes out of the linear density field.

Our algorithm can also be seen as a specific example of a bias model which is local in Lagrangian space. This is expected to have relevant consequences on galaxy clustering. Because of this local Lagrangian character, our model differs strongly from the local Eulerian bias prescription applied by Fry & Gaztañaga (1993) to the analysis of the hierarchical correlation functions. A simple test of our theory can be obtained by analysing the behaviour of the bispectrum (or the skewness), whose shape (scale) dependence will be shown to be directly sensitive to the assumption of local bias in Lagrangian versus Eulerian space.

Our results for the evolved halo distribution generally allow us to study their statistical properties at the required level of non-linearity, and could be further implemented to generate mock halo catalogues, starting from low-resolution numerical

simulations of the dissipationless matter component. These results have important implications for the study of the redshift evolution of galaxy clustering, a problem made of compelling relevance by the growing body of observational data at high redshift which are being produced by the new generation of large telescopes. A general study of this problem has been recently performed by Peacock (1997) and Matarrese et al. (1997); the latter pointed out that knowledge of the evolution of the effective bias for the various classes of objects is a key ingredient in the comparison of theoretical scenarios of structure formation with observational data on clustering at high redshift. Kauffmann, Nusser & Steinmetz (1997) used both semi-analytical methods and N-body techniques to study the physical origin of bias in galaxies of different luminosity and morphology.

The plan of the paper is as follows. In Section 2 we define our halo counting field, within the linear approximation, both in the Lagrangian and Eulerian context. The non-linear evolution of the halo clustering is analysed in Section 3, where we also compute the bispectrum and skewness of the evolved halo distribution. Section 4 contains a general discussion of our results and some conclusions.

2 STOCHASTIC APPROACH TO HALO COUNTING AND CLUSTERING

2.1 Basic tools and notation

Let us assume that the mass density contrast $\epsilon(\boldsymbol{q})$, linearly extrapolated to the present time, is a statistically homogeneous and isotropic Gaussian random field completely determined by its power spectrum $P(k)$. Here $\boldsymbol{q}$ represents a comoving Lagrangian coordinate. A smoothed version of the field $\epsilon(\boldsymbol{q})$ is obtained by convolving it with a rotationally invariant filter $W_R(q)$, containing a resolution scale R, with associated mass $M \sim \rho_b R^3$, ρ_b being the background mean density at $z=0$,

$$\epsilon_R(\boldsymbol{q}) = \int \mathrm{d}\boldsymbol{q}'\, W_R(|\boldsymbol{q}-\boldsymbol{q}'|)\,\epsilon(\boldsymbol{q}') \equiv \epsilon_M(\boldsymbol{q}). \tag{1}$$

The smoothed field is also Gaussian with one-point distribution function $G_{\sigma_M}(\epsilon_M) = (2\pi\sigma_M^2)^{-1/2}\exp(-\epsilon_M^2/2\sigma_M^2)$, where σ_M^2 denotes the variance of ϵ_M, $\sigma_M^2 \equiv \langle \epsilon_M^2 \rangle = (2\pi^2)^{-1}\int_0^\infty \mathrm{d}k\, k^2 P(k)\,\tilde{W}(kR)^2$. The symbol $\tilde{W}(kR)$ indicates the Fourier transform of the filter function. In the following, we will often be concerned with the joint distribution of the fields $\epsilon_{M_1}(\boldsymbol{q})$ and $\epsilon_{M_2}(\boldsymbol{q})$. The two-point correlation function of the linear density contrast smoothed on the scale R_1 and R_2 is

$$\xi_{12}(q) = \langle \epsilon_{M_1}(\boldsymbol{q}_1)\,\epsilon_{M_2}(\boldsymbol{q}_2) \rangle = \frac{1}{2\pi^2}\int_0^\infty \mathrm{d}k\, k^2 P(k)\,\tilde{W}(kR_1)\,\tilde{W}(kR_2)\,j_0(kq), \tag{2}$$

where $q = |\boldsymbol{q}_1 - \boldsymbol{q}_2|$, and $j_0(x)$ is the spherical Bessel function of order zero. We term σ_{12}^2 the value assumed by ξ_{12} in the limit $q \to 0$.

The properties of the filtered quantities clearly depend upon the choice of the window function. For instance, the relation between the mass enclosed by a top-hat filter $W_R(q) = 3\Theta(R-q)/4\pi R^3$ [where $\Theta(x)$ is the Heaviside step function] is the standard $M(R) = 4\pi\rho_b R^3/3$. Instead, for a Gaussian window, $W_R(q) = (2\pi R^2)^{-3/2}\exp(-q^2/2R^2)$, the enclosed mass is $M(R) = (2\pi)^{3/2}\rho_b R^3$. These two masses coincide for $R_G = 0.64 R_{TH}$ (Bardeen et al. 1986).

In the literature, the sharp top-hat filtering has been alternatively adopted in Fourier space, $\tilde{W}_R(k) = \Theta(k_R - k)$, where $k_R = 1/R$ and $k = |\boldsymbol{k}|$. The most remarkable property of this filter is that each decrease of the smoothing radius adds up a new set of uncorrelated modes (Bardeen et al. 1986; Bond et al. 1991; Lacey & Cole 1993). This also implies that, for example, the correlation function in equation (2) simplifies to $\xi_{12} = \xi_{11}$, whenever $k_{R_1} < k_{R_2}$; consequently, $\sigma_{12} = \sigma_{11} \equiv \sigma_1$. In practice, the information is always erased below the larger of the two smoothing lengths. This property will be particularly useful in the next sections. For this 'sharp k-space' filter, the main difficulty is how to associate a mass $M(R)$ to the cut-off wavenumber k_R. Lacey & Cole (1993) give the expression $M(R) = 6\pi^2\rho_b k_R^{-3}$, which coincides with the mass within a top-hat filter if one takes $k_R = 2.42/R_{TH}$.

In the next section we introduce the halo counting random fields that allow a fully stochastic description of the biased haloes distribution. To illustrate how our formalism works, we first show how to derive the PS mass function by performing a simple averaging of our stochastic counts.

2.2 Lagrangian mass function: Press–Schechter theory

PS proposed a simple model to compute the comoving number density of collapsed haloes directly from the statistical properties of the linear density field, assumed to be Gaussian. According to the PS theory, a patch of fluid is part of a collapsed region of scale larger than $M(R)$ if the value of the smoothed linear density contrast on that scale exceeds a suitable threshold t_f. The idea is to use a global threshold in order to mimic non-linear dynamical effects ending up with halo collapse and virialization. An exact value for t_f can be obtained by describing the evolution of the density perturbations according to the spherical top-hat model. In this case, a fluctuation of amplitude ϵ will collapse at the redshift z_f such that $\epsilon(\boldsymbol{q}) = t_f \equiv \delta_c(z_f)$,

where $D(z)$ denotes the linear growth factor of density perturbations normalized as $D(0)=1$. In the Einstein–de Sitter universe and during the matter-dominated era the critical value δ_c does not depend on any cosmological parameter and is given by $\delta_c=3(12\pi)^{2/3}/20\simeq 1.686$, while, for general non-flat geometries, its value shows a weak dependence on the density parameter Ω, the cosmological constant Λ and the Hubble constant H (e.g. Lacey & Cole 1993), thus on redshift. In a flat universe with vanishing cosmological constant $D(z)=(1+z)^{-1}$; explicit expressions for the linear growth factor are given in Appendix A for general Friedmann models.

A local version of the PS approach can be built up in terms of stochastic counting operators acting on the underlying Gaussian density field, as follows. The number of haloes per unit mass, contained in the unit comoving volume centred in $\boldsymbol{q}$, identified by the collapse threshold $t_f(z_f)$, is described as a density field of a point process by

$$\mathcal{N}_h^L(\boldsymbol{q}|M, t_f) = -2\,\frac{\rho_b}{M}\,\frac{\partial}{\partial M}\,\Theta[\epsilon_M(\boldsymbol{q}) - t_f]. \tag{3}$$

Note that the quantity $\mathcal{N}_h^L(\boldsymbol{q}|M, t_f)$ is non-zero only when the filtered density contrast in $\boldsymbol{q}$ upcrosses, or downcrosses, the threshold t_f, by varying the smoothing length R (or the corresponding mass M). The factor of 2, appearing in the expression of $\mathcal{N}_h^L(\boldsymbol{q}|M, t_f)$, is needed in order to obtain the right normalization of the mass function, in which case it has been shown to be intimately related to the solution of the cloud-in-cloud problem (Peacock & Heavens 1990; Bond et al. 1991; Cole 1991), at least for sharp k-space filtering. At this level, our description should be thought of as a sort of differential version of Kaiser's bias model (Kaiser 1984), that defines a population of objects with the right average halo abundance and their related clustering properties, rather than a detailed modelling of how structures form from the primordial density field. In a forthcoming paper, however, we will show that the present approach is fully consistent with a rigorous treatment of the cloud-in-cloud problem (Porciani et al. 1998). In that approach, halo correlations will be obtained from pairs of first-upcrossing 'times' for spatially correlated random walks above the collapse threshold t_f.

It can be seen from equation (3) that a population of haloes is uniquely specified by the two parameters M and t_f. In the standard PS formulation, t_f is interpreted as a sort of time variable, related to the formation redshift z_f, which decreases with real time, as every halo continuously accretes matter. In this sense one can say that, for a continuous density field with infinite mass resolution, each halo disappears as soon as it forms to originate another halo of larger mass.

Alternatively, instead of considering t_f as a time variable, one can use it simply as a label attached to each halo. The haloes so labelled can be thought as keeping their identity during the subsequent evolution at any observation redshift z. This is not in contrast with the fact that in the real Universe dark matter haloes undergo merging at some finite rate (e.g. Lacey & Cole 1993, 1994). Within such a picture, in fact, the physical processes of accretion and merging reduce to the trivial statement that haloes identified by a given threshold are necessarily included in catalogues of lower threshold, so that, in the limit of infinite mass resolution, only haloes with $z_f=z$ would actually survive. Nevertheless, keeping z_f distinct for z may have several advantages, among which the possibility of allowing for a more realistic description of galaxy and cluster formation inside haloes, for both the evolution of the luminosity function (Cavaliere, Colafrancesco & Menci 1993; Manrique & Salvador–Solé 1996) and of the galaxy bias (e.g. MW; Kauffmann et al. 1997). Let us stress, however, that we are not addressing here the issue of galaxy or cluster merging: our method is completely general in this respect and allows us to span all possible models, from the instantaneous merging hypothesis ($z_f=z$) to the case of no merging at all (z_f fixed for changing $z\le z_f$).

In what follows, therefore, we will assume that we can deal with the halo population specified by the formation threshold t_f at any redshift z. Only in this sense we will say that we 'ignore' the effects of merging in our description: merging can be exactly recovered at any step, and with any assumed mass resolution, as the relation between z_f and z. To implement this idea it is enough to scale appropriately the argument of the Heaviside function in equation (3), which can be recast in the form

$$\mathcal{N}_h^L(\boldsymbol{q}, z|M, z_f) = -2\,\frac{\rho_b}{M}\,\frac{\partial}{\partial M}\,\Theta[\epsilon_M(\boldsymbol{q}, z) - \delta_f(z, z_f)], \tag{4}$$

where $\epsilon_M(\boldsymbol{q}, z)\equiv D(z)\,\epsilon(\boldsymbol{q})$, and $\delta_f(z, z_f)\equiv\delta_c D(z)/D(z_f)$. It can be easily seen that the ensemble average of the counting field $\mathcal{N}_h^L(\boldsymbol{q}, z|M, z_f)$ corresponds to the PS mass function

$$\langle\mathcal{N}_h^L(\boldsymbol{q}, z|M, z_f)\rangle\,\mathrm{d}M = n_{PS}(z|M, z_f)\,\mathrm{d}M, \tag{5}$$

where

$$n_{PS}(z|M, z_f)\,\mathrm{d}M \equiv \frac{1}{\sqrt{2\pi}}\,\frac{\rho_b}{M}\,\frac{\delta_f(z, z_f)}{\sigma_M^3(z)}\exp\left[-\frac{\delta_f^2(z, z_f)}{2\sigma_M^2(z)}\right]\left|\frac{\mathrm{d}\sigma_M^2(z)}{\mathrm{d}M}\right|\mathrm{d}M. \tag{6}$$

Note that we emphasized the z-dependence of the comoving mass function, although it is straightforward to verify that the value of $n_{PS}(z|M, z_f)$ does not change with z. In fact, since we are ignoring the effects of merging, once a class of haloes has

been identified, their mean comoving density remains constant in time. Thus, as far as the mass function is concerned, the introduction of the observation redshift z is somewhat more formal than physical. However, this distinction will be far more significant in the next sections, where, in order to compute the halo-to-mass bias factor, we will relate the Lagrangian distribution of a population of haloes selected at z_f to the mass density fluctuation field linearly extrapolated to the redshift z. Models of galaxy formation which assume that galaxies form at a given redshift z_f with some initial bias factor and that their subsequent motion is purely caused by gravity (e.g. Dekel 1986; Dekel & Rees 1987; Nusser & Davis 1994; Fry 1996) can be easily accommodated into this scheme.

To conclude this section, let us consider the integral stochastic process

$$\int_M^\infty \mathrm{d}M' M' \mathcal{N}_h^L(\boldsymbol{q}, z|M', z_f) = 2\rho_b \Theta[\epsilon_M(\boldsymbol{q}, z) - \delta_f(z, z_f)], \tag{7}$$

representing the fraction of mass, in the unit Lagrangian comoving volume centred in $\boldsymbol{q}$, which at redshift z_f has formed haloes more massive than M. This coincides with the original Kaiser bias model (Kaiser 1984) up to the multiplicative factor $2\rho_b$, which is irrelevant for calculating correlation function.

2.3 Conditional Lagrangian mass function

The PS theory reviewed in the previous section describes the overall distribution of halo masses in a homogeneous universe of mean density ρ_b. However, of cosmological interest is also, for instance, the estimate of the halo distribution within rich or poor environments (which can be related to the galaxy number enhancement per unit mass in rich clusters or in voids), thus justifying the investigation of the distribution of halo masses conditioned to lie within a larger uncollapsed container of given density. The conditional mass function has been studied by several authors (e.g. Bond et al. 1991; Bower 1991; Lacey & Cole 1993).

We extend here the approach introduced in the previous section in order to derive the conditional mass function. Specifically, we calculate the comoving mass function, in the mass range M to $M + \mathrm{d}M$, for objects contained in a large region of dimension R_0, corresponding to a mass M_0, with local density contrast $\epsilon_0 \equiv \epsilon_{M_0}$. We will require $\epsilon_0 \ll \delta_f$ and $R_0 \gg R$, to ensure that the container is not collapsed yet by the epoch z_f, and that it encloses a non-negligible population of objects.

In order to mimic these environmental effects, we modify the halo counting field according to

$$\mathcal{N}_h^L(\boldsymbol{q}, z|M, z_f|M_0, \epsilon_0) = -\frac{2}{N_0}\frac{\rho_b}{M}\frac{\partial}{\partial M}\Theta[\epsilon_M(\boldsymbol{q}, z) - \delta_f(z, z_f)]\,\delta_D[\epsilon_{M_0}(\boldsymbol{q}, z) - \epsilon_0], \tag{8}$$

where δ_D denotes the Dirac delta function, and $N_0 \equiv \langle \delta_D[\epsilon_{M_0}(\boldsymbol{q}, z) - \epsilon_0] \rangle$ is the normalization constant. Here the scalar ϵ_0 indicates the value of the random field $\epsilon_{M_0}(\boldsymbol{q}, z)$. Taking the ensemble average (and using the cross-variance σ_{ij} for a sharp k-space filter), one eventually obtains

$$\langle \mathcal{N}_h^L(\boldsymbol{q}, z|M, z_f|M_0, \epsilon_0) \rangle\, \mathrm{d}M = n_{PS}(z|M, z_f|M_0, \epsilon_0)\, \mathrm{d}M, \tag{9}$$

where

$$n_{PS}(z|M, z_f|M_0, \epsilon_0)\, \mathrm{d}M = \frac{1}{\sqrt{2\pi}}\frac{\rho_b}{M}\frac{\delta_f(z, z_f) - \epsilon_0}{[\sigma_M^2(z) - \sigma_0^2(z)]^{3/2}} \exp\left\{-\frac{[\delta_f(z, z_f) - \epsilon_0]^2}{2[\sigma_M^2(z) - \sigma_0^2(z)]}\right\}\left|\frac{\mathrm{d}\sigma_M^2(z)}{\mathrm{d}M}\right| \mathrm{d}M. \tag{10}$$

This straightforward calculation shows how to obtain results already known in the literature by simply starting from the random field in equation (8); averaging that halo counting field leads to the expected conditional mass function. However, unlike previous treatments, once the halo counting field has been consistently defined, other statistics, like the two-point halo correlation function, can be calculated. We will carry out this programme in the next section.

2.4 Lagrangian clustering: halo-to-mass bias from correlations

In this section we will compute the halo–halo correlation function, which coincides with the correlation function of our random counting field. Specifically, we will calculate the Lagrangian halo correlation function from the Lagrangian counting field $\mathcal{N}_h^L(\boldsymbol{q}, z|M, z_f)$. By definition, the correlation function of this stochastic process is given by

$$\xi_{hh}^L(q, z|M_1, z_1; M_2, z_2) = \frac{\langle \mathcal{N}_h^L[\boldsymbol{q}_1, z|M_1, \delta_f(z, z_1)]\, \mathcal{N}_h^L(\boldsymbol{q}_2, z|M_2, \delta_f(z, z_2)] \rangle}{\langle \mathcal{N}_h^L[\boldsymbol{q}_1, z|M_1, \delta_f(z, z_1)] \rangle \langle \mathcal{N}_h^L[\boldsymbol{q}_2, z|M_2, \delta_f(z, z_2)] \rangle} - 1, \tag{11}$$

where $q=|\boldsymbol{q}_1-\boldsymbol{q}_2|$. Performing the ensemble average over the Gaussian fields $\epsilon_{M_1}(\boldsymbol{q})$ and $\epsilon_{M_2}(\boldsymbol{q})$, we obtain

$$\langle \mathcal{N}_{\rm h}^{\rm L}[\boldsymbol{q}_1, z|M_1, \delta_{\rm f}(z, z_1)]\,\mathcal{N}_{\rm h}^{\rm L}[\boldsymbol{q}_2, z|M_2, \delta_{\rm f}(z, z_2)]\rangle = \frac{4\rho_{\rm b}^2}{M_1 M_2}\,\frac{\partial}{\partial M_1}\,\frac{\partial}{\partial M_2}\int_{\delta_{\rm f}(z,z_1)}^{\infty}\int_{\delta_{\rm f}(z,z_2)}^{\infty} {\rm d}\alpha_1\,{\rm d}\alpha_2 G_2(\alpha_1, \alpha_2), \tag{12}$$

where $G_2(\alpha_1, \alpha_2)$ denotes the bivariate Gaussian distribution

$$G_2(\alpha_1, \alpha_2) = (2\pi\sigma_1\sigma_2\sqrt{1-\omega^2})^{-1}\exp\left[-\left(\frac{\alpha_1^2}{\sigma_1^2}+\frac{\alpha_2^2}{\sigma_2^2}-2\omega\,\frac{\alpha_1}{\sigma_1}\,\frac{\alpha_2}{\sigma_2}\right)\Big/ 2(1-\omega^2)\right], \tag{13}$$

with normalized correlation $\omega(q)=\xi_{12}(q)/\sigma_{M_1}\sigma_{M_2}$ and $\sigma_i \equiv D(z)\,\sigma_{M_i}$.

The full exact expression for the halo–halo correlation function can be obtained after an incredibly long algebraic computation. We report here only the final expression. Defining $\delta_{{\rm f}i}\equiv\delta_{\rm f}(z, z_i)$, we have

$$\begin{aligned}1+\xi_{\rm hh}^{\rm L}(q, z|M_1, z_1; M_2, z_2) = {} & \frac{1}{\sqrt{1-\omega^2}}\left\{\frac{{\rm d}\sigma_1}{{\rm d}M_1}\,\frac{{\rm d}\sigma_2}{{\rm d}M_2}+\frac{\sigma_2^2}{\delta_{\rm f2}(1-\omega^2)}\left(\frac{\delta_{\rm f1}}{\sigma_1}-\omega\,\frac{\delta_{\rm f2}}{\sigma_2}\right)\frac{{\rm d}\sigma_1}{{\rm d}M_1}\,\frac{\partial\omega}{\partial M_2}\right.\\ & +\frac{\sigma_1^2}{\delta_{\rm f1}(1-\omega^2)}\left(\frac{\delta_{\rm f2}}{\sigma_2}-\omega\,\frac{\delta_{\rm f1}}{\sigma_1}\right)\frac{\partial\omega}{\partial M_1}\,\frac{{\rm d}\sigma_2}{{\rm d}M_2}+\frac{\sigma_1^2\sigma_2^2}{\delta_{\rm f1}\delta_{\rm f2}}\,\frac{\partial^2\omega}{\partial M_1\partial M_2}\\ & +\frac{\sigma_1^2\sigma_2^2}{\delta_{\rm f1}\delta_{\rm f2}(1-\omega^2)^2}\left[\omega(1-\omega^2)+(1+\omega^2)\,\frac{\delta_{\rm f1}}{\sigma_1}\,\frac{\delta_{\rm f2}}{\sigma_2}-\omega\left(\frac{\delta_{\rm f1}^2}{\sigma_1^2}+\frac{\delta_{\rm f2}^2}{\sigma_1^2}\right)\right]\\ & \left.\times\frac{\partial\omega}{\partial M_1}\,\frac{\partial\omega}{\partial M_2}\right\}\exp\left[-\frac{\omega^2\left(\frac{\delta_{\rm f1}^2}{\sigma_1^2}+\frac{\delta_{\rm f2}^2}{\sigma_2^2}\right)-2\omega\,\frac{\delta_{\rm f1}}{\sigma_1}\,\frac{\delta_{\rm f2}}{\sigma_2}}{2(1-\omega^2)}\right]\left(\frac{{\rm d}\sigma_1}{{\rm d}M_1}\,\frac{{\rm d}\sigma_2}{{\rm d}M_2}\right)^{-1}.\end{aligned} \tag{14}$$

This expression can be easily shown to be independent of the observation redshift z. A remark is now appropriate. Our formalism describes the halo distribution as a discrete-point process. However, actual haloes are extended in size. This is why, as also seen in numerical simulations, for separation smaller than the typical Lagrangian radius of the halo, the correlation function abruptly reaches the value -1: a sort of 'exclusion principle' for extended haloes. Thus we expect that the correlation function in equation (14) can be a reliable description of halo clustering only for $q\gtrsim\max(R_1, R_2)$. Another point concerns the use of finite mass resolution as in N-body simulations. The proper analytical correlation to compare with in that case is the integral of $\xi_{\rm hh}^{\rm L}\,n_{\rm PS}(M_1)\,n_{\rm PS}(M_2)$ over the specified mass interval, appropriately normalized.

Since the action of the window functions on the correlations is negligible for lags q much larger than the smoothing lengths, $q\gg R_1$ and $q\gg R_2$, for the normalized correlation we obtain $\omega(q)\simeq\xi_{\rm m}(q)/\sigma_{M_1}\sigma_{M_2}$ [where $\xi_{\rm m}(q)$ is the linear mass autocorrelation function] and, eventually, for the halo correlation

$$\xi_{\rm hh}^{\rm L}(q, z|M_1, z_1; M_2, z_2) = b_1^{\rm L}(z|M_1, z_1)\,b_1^{\rm L}(z|M_2, z_2)\,\xi_{\rm m}(q, z)+\frac{1}{2}\,b_2^{\rm L}(z|M_1, z_1)\,b_2^{\rm L}(z|M_2, z_2)\,\xi_{\rm m}^2(q, z)+\ldots \tag{15}$$

Explicitly, the first two bias parameters read

$$b_1^{\rm L}(z|M, z_{\rm f}) = \frac{\delta_{\rm f}(z, z_{\rm f})}{\sigma_M^2(z)}-\frac{1}{\delta_{\rm f}(z, z_{\rm f})} = \frac{D(z_{\rm f})}{D(z)}\left[\frac{\delta_{\rm c}}{D(z_{\rm f})^2\sigma_M^2}-\frac{1}{\delta_{\rm c}}\right], \tag{16}$$

$$b_2^{\rm L}(z|M, z_{\rm f}) = \frac{1}{\sigma_M^2(z)}\left[\frac{\delta_{\rm f}^2(z, z_{\rm f})}{\sigma_M^2(z)}-3\right] = \frac{1}{D(z)^2\sigma_M^2}\left[\frac{\delta_{\rm c}^2}{D(z_{\rm f})^2\sigma_M^2}-3\right]. \tag{17}$$

These expressions for the bias factors generalize, in a sense, those concerning the clustering properties of dark matter haloes in Lagrangian space obtained by MW and Mo et al. (1996), with the relevant difference that we have obtained the bias factor from the behaviour of the halo two-point correlation function. Also relevant is the fact that, unlike MW, we obtained our Lagrangian correlation function without introducing any background scale R_0, which allows us to extend its validity down to spatial separation comparable to the halo size $R\ll R_0$. A calculation of the leading behaviour of the correlation deriving from equations (11) and (12) has been already attempted by Kashlinsky (1987) who, however, missed the contributions originated by the differentiation of ω with respect to M_1 and M_2, thereby obtaining an incomplete expression for $b_1^{\rm L}$.

The halo correlation function in Lagrangian space, $\xi^{\rm L}_{\rm hh}$ from equation (14), with $M_1=M_2=M$ and $z_1=z_2\equiv z_{\rm f}$, is shown in Fig. 1 for two scale-free power spectra, $P(k)\propto k^n$, with spectral index $n=-2$ and -1, in the Einstein–de Sitter case. The two-point function is calculated for various halo masses in units of the characteristic mass, M_*, defined so that $\sigma_{M_*}\equiv t_{\rm f}=\delta_{\rm c}/D(z_{\rm f})$, with top-hat filtering;[1] the spatial dependence is shown as a function of the scaling variable q/R, which eliminates any residual redshift dependence. Also shown is the mass autocorrelation function $\xi_{\rm m}$ and an estimate of the Lagrangian halo two-point function obtained as $(b^{\rm L}_1)^2\xi_{\rm m}$, for $M\neq M_*$, and $(b^{\rm L}_2)^2\xi^2_{\rm m}/2$, for $M=M_*$, as in this case $b^{\rm L}_1=0$. Note that such an estimate of $\xi^{\rm L}_{\rm hh}$ always provides an accurate fit to its analytical expression for separation a few times larger than the halo size. The characteristic behaviour of the halo correlation function for $M=M_*$, where the linear bias vanishes, is actually a peculiarity of the Lagrangian case (see also MW). As we will see below, the Eulerian halo correlation function does not show such a drastic change of slope in the same mass range.

2.5 Peak–background split

In the previous paragraphs we computed number densities and correlation functions of haloes in Lagrangian space. However, after their identification, these haloes in embryo move following the gravitational field, modifying their original spatial distribution. One issue to address is how, for instance, the conditional halo number density per unit mass changes as a consequence of gravitational evolution. Furthermore, of interest is to quantify the evolution of clustering in terms of the halo correlation functions, or in terms of the halo-to-mass bias. Both problems can be dealt with by defining Eulerian halo counting fields, in the same spirit as we did for the Lagrangian case.

Essentially, our approach to the clustering evolution can be based on a generalization of the so-called peak–background split, first proposed by Bardeen et al. (1986), which basically consists in splitting the mass perturbations in fine-grained (*peak*) and coarse-grained (*background*) components.[2] The underlying idea is to ascribe the collapse of objects on small scales to the high-frequency modes of the density field, while the action of large-scale structures on these non-linear condensations is due to the remaining low-frequency modes. At the linear level the resulting effect of these long wavelengths is simply modelled as a shift of the local background density.

In the spirit of the peak–background split, we define the linear density field smoothed on a given scale ϵ_M as consisting of two complementary and superimposed components, namely $\epsilon_M=\epsilon_{\rm bg}+\epsilon_{\rm pk}$. Adopting as window function the sharp k-space filter, we define as 'background' component the density contrast smoothed on the scale $R_0=1/k_0$:

$$\epsilon_{\rm bg}(\boldsymbol{q},z)\equiv\int\frac{{\rm d}\boldsymbol{k}}{(2\pi)^3}\,\tilde{\epsilon}(\boldsymbol{k},z)\,\Theta(k_0-k)\,{\rm e}^{{\rm i}\boldsymbol{k}\cdot\boldsymbol{q}}. \tag{18}$$

The 'peak' component is instead obtained by smoothing the mass density fluctuation with the band-pass filter $\Theta(k_M-k)\,\Theta(k-k_0)$, namely

$$\epsilon_{\rm pk}(\boldsymbol{q},z)\equiv\int\frac{{\rm d}\boldsymbol{k}}{(2\pi)^3}\,\tilde{\epsilon}(\boldsymbol{k},z)\,\Theta(k_M-k)\,\Theta(k-k_0)\,{\rm e}^{{\rm i}\boldsymbol{k}\cdot\boldsymbol{q}}, \tag{19}$$

where $k_M=1/R_M$, with $M\propto\rho_{\rm b}R^3_M$ and $M_0\propto\rho_{\rm b}R^3_0$ the masses enclosed by the two filters. So, the peak component contains only modes with wavenumber in the interval $[k_0,k_M]$. Note that in the linear regime, with Gaussian initial conditions, the peak and background components are statistically independent, i.e.,

$$\langle\epsilon_{\rm pk}(\boldsymbol{q}_1,z)\,\epsilon_{\rm bg}(\boldsymbol{q}_2,z)\rangle=0, \tag{20}$$

for, by construction, the two fields do not share any common Fourier mode. To summarize: provided that the collapsed object is described according to the spherical model, as in the PS theory, the peak field $\epsilon_{\rm pk}(\boldsymbol{q},z)$ can be thought as evolving in a local environment with effective mean density $\rho_{\rm b}[1+\epsilon_{\rm bg}(\boldsymbol{q},z)]$. This implies that the collapse condition can be written as $\epsilon_{\rm pk}(\boldsymbol{q},z)=\delta_{\rm f}(z,z_{\rm f})-\epsilon_{\rm bg}(\boldsymbol{q},z)$.

2.6 Eulerian halo counting field and bias

The previous analysis shows how the PS and the conditional Lagrangian mass functions can be obtained by averaging properly defined halo counting random fields. It is thus legitimate to explore the possibility of building up analogous counting processes in the Eulerian world. Our approach here will be based on the peak–background split technique described above.

[1]We are adopting here the MW definition of M_*, which, though differing from the standard PS one, $\sigma_{M_*}\equiv t_{\rm f}/\sqrt{2}$, is more convenient for our present purposes.

[2]We are here making a rather liberal use of the work 'peak', to mean the fine-grained component of the linear density field.

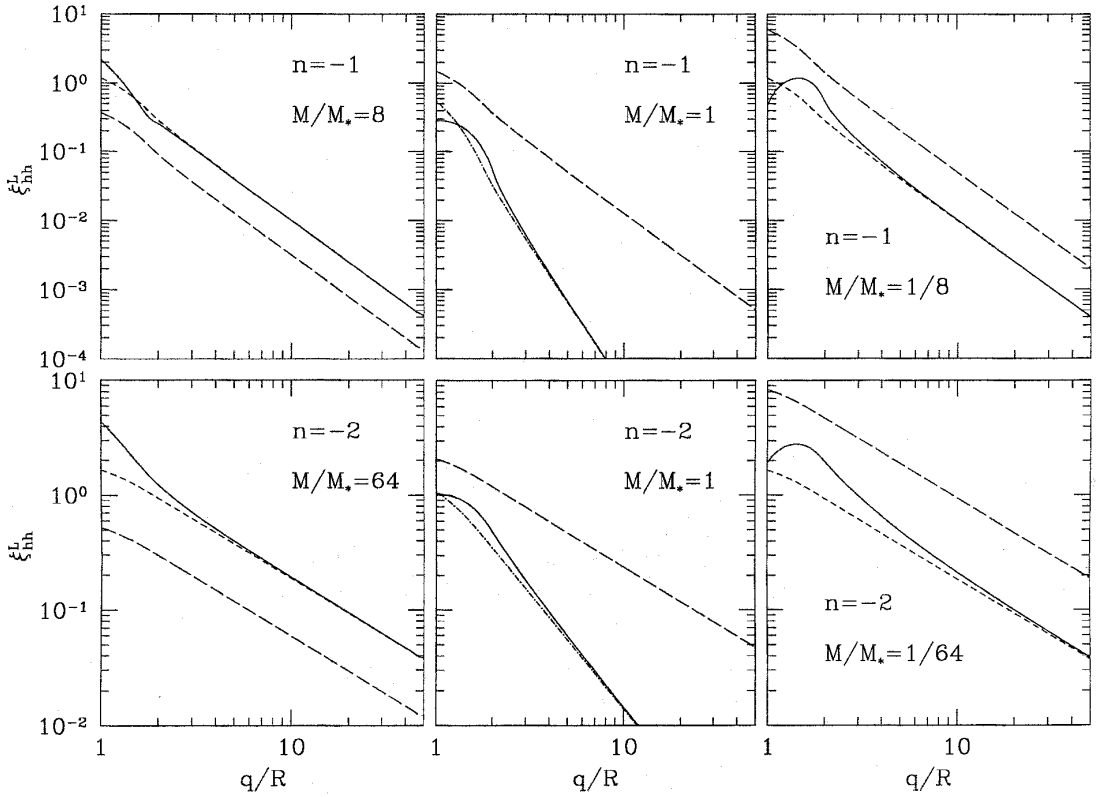

Figure 1. The exact Lagrangian halo correlation function in an Einstein–de Sitter universe (solid lines) is shown for scale-free models with spectral index $n=-1$ and -2, and for various masses. In each panel we set $M_1=M_2=M$ and $z_1=z_2$. Results are plotted in terms of the scaling variables M/M_* and q/R (with R the top-hat radius corresponding to the halo mass M), which makes the resulting curves redshift-independent. For comparison, the linear mass autocorrelation function smoothed on the halo scale is also shown with long-dashed lines. The short-dashed lines represent the linear bias approximation for the halo correlations: $(b_1^{\rm L})^2\xi_{\rm m}$. In the central panels, where $b_1^{\rm L}=0$, the dot–dashed lines show, instead, the second-order approximnation for $\xi^{\rm L}_{\rm hh}$. Each column contains panels that refer to the same mass variance ($\sigma_M^2/t_{\rm f}^2=1/4$, 1, 4) and so the same Lagrangian bias factors. Notice that, for separation a few times the halo size, the first non-vanishing term of equation (15) always gives an accurate approximation to the exact halo correlation. This implies that, for M_* objects, $\xi^{\rm L}_{\rm hh}\approx(b_2^{\rm L})^2\xi_{\rm m}^2/2$.

Let us define the Eulerian counting field of haloes collapsed at redshift $z_{\rm f}$ and observed at z as

$$\mathcal{N}_{\rm h}^{\rm E}(\boldsymbol{q}, z|M, z_{\rm f})\equiv[1+\epsilon_{\rm bg}(\boldsymbol{q}, z)]\,\mathcal{N}_{\rm h}^{\rm L}(\boldsymbol{q}, z|M, z_{\rm f})=-\frac{2\rho_{\rm b}}{M}\,[1+\epsilon_{\rm bg}(\boldsymbol{q}, z)]\,\frac{\partial}{\partial M}\,\Theta\{\epsilon_{\rm pk}(\boldsymbol{q}, z)-[\delta_{\rm f}(z, z_{\rm f})-\epsilon_{\rm bg}(\boldsymbol{q}, z)]\}. \tag{21}$$

The watchful reader might wonder about our use of the Lagrangian variable $\boldsymbol{q}$ within the Eulerian framework; however, in linear theory, $\boldsymbol{x}=\boldsymbol{q}$. Once again, the redshift z must be thought of as the redshift the sampled objects have at the epoch of observation. It is worth noting that equation (21) is fully consistent with the analysis in Cole & Kaiser (1989). Most importantly, our treatment allows for a local description. Let us stress here that the factor $(1+\epsilon_{\rm bg})$, connecting the Eulerian to the Lagrangian counting field, simply comes from mass conservation in Eulerian space (see also Section 3.1 and, in particular, equation 38); this point has been discussed in more detail by Kofman et al. (1994).

Now consider the integral stochastic process

$$\int_M^\infty {\rm d}M'\,M'\,\mathcal{N}_{\rm h}^{\rm E}(\boldsymbol{q}, z|M', z_{\rm f})=2\rho_{\rm b}[1+\epsilon_{\rm bg}(\boldsymbol{q}, z)]\,\Theta[\epsilon_M(\boldsymbol{q}, z)-\delta_{\rm f}(z, z_{\rm f})]; \tag{22}$$

this represents the fraction of mass, in the unit Eulerian comoving volume centred in $\boldsymbol{q}$, which at redshift $z_{\rm f}$ will form haloes more massive than M. For $M_0\to M$, $\epsilon_{\rm bg}\to\epsilon_M$ and the above relation coincides (up to the usual fudge factor of 2, having no effect on correlations) with the *weighted bias* model of Catelan et al. (1994). An extended version of this scheme, called 'censoring bias', has been recently proposed by Mann, Peacock & Heavens (1998). Thus the weighted bias is just the Eulerian version, within linear theory, of the Kaiser (1984) bias model.

Of course, further specifications could be added to our Eulerian counting field. For instance, we might ask that the background scale has not yet collapsed by the epoch $z_{\rm f}$; in such a case we should multiply the above stochastic process by the

factor $\Theta[\delta_{\rm f}(z, z_{\rm f}) - \epsilon_{\rm bg}(\boldsymbol{q}, z)]$. Extra details of this kind would, however, make negligible changes to our final results, provided $\sigma_M^2 \gg \sigma_0^2$.

As in the Lagrangian case, to calculate the mean halo number density per unit mass, one needs to ensemble average $\mathcal{N}_{\rm h}^{\rm E}(\boldsymbol{q}, z|M, z_{\rm f})$. Let us analyse this operation in more detail. Because of the way the Eulerian counting process has been defined, it is clear that $\mathcal{N}_{\rm h}^{\rm E}(\boldsymbol{q}, z|M, z_{\rm f})$ depends on two random fields, specifically $\epsilon_{\rm bg}$ and $\epsilon_{\rm pk}$. So the ensemble average $\langle \mathcal{N}_{\rm h}^{\rm E} \rangle$ can be interpreted as a double average over these fields, i.e., $\langle \mathcal{N}_{\rm h}^{\rm E} \rangle \equiv \langle \langle \mathcal{N}_{\rm h}^{\rm E} \rangle_{\epsilon_{\rm pk}} \rangle_{\epsilon_{\rm bg}}$. The statistics of the field $\mathcal{N}_{\rm h}^{\rm E}$ can be described in terms of nth-order correlation functions, $\langle \langle \mathcal{N}_{\rm h}^{\rm E}(\boldsymbol{q}_1) \dots \mathcal{N}_{\rm h}^{\rm E}(\boldsymbol{q}_n) \rangle_{\epsilon_{\rm pk}} \rangle_{\epsilon_{\rm bg}}$. The exact calculation of these quantities is rather difficult. However, because of the short-scale coherence of the peak field, implied by the 'infrared' cut-off at k_0, its covariance $\langle \epsilon_{\rm pk}(\boldsymbol{q}_i)\, \epsilon_{\rm pk}(\boldsymbol{q}_i + \boldsymbol{r}) \rangle_{\epsilon_{\rm pk}}$ vanishes whenever $r \gg R_0$, so that we can simplify the general halo correlations above as $\langle \langle \mathcal{N}_{\rm h}^{\rm E}(\boldsymbol{q}_1) \dots \mathcal{N}_{\rm h}^{\rm E}(\boldsymbol{q}_n) \rangle_{\epsilon_{\rm pk}} \rangle_{\epsilon_{\rm bg}} \approx \langle \langle \mathcal{N}_{\rm h}^{\rm E}(\boldsymbol{q}_1) \rangle_{\epsilon_{\rm pk}} \dots \langle \mathcal{N}^{\rm E}(\boldsymbol{q}_n) \rangle_{\epsilon_{\rm pk}} \rangle_{\epsilon_{\rm bg}}$, provided that we consider sets of points $\boldsymbol{q}_i$, $i = 1, \dots, N$, with relative separation $r_{ij} \equiv |\boldsymbol{q}_i - \boldsymbol{q}_j| \gg R_0$. Therefore, with the purpose of calculating the mean Eulerian halo number density per unit mass and Eulerian halo correlations, we can make the replacement $\mathcal{N}_{\rm h}^{\rm E} \rightarrow \langle \mathcal{N}_{\rm h}^{\rm E} \rangle_{\epsilon_{\rm pk}} \equiv N_{\rm h}^{\rm E}$, with only negligible loss of accuracy. According to the definition of $\mathcal{N}_{\rm h}^{\rm E}$ in equation (21), the latter ensemble average gives

$$N_{\rm h}^{\rm E}(\boldsymbol{q}, z|M, z_{\rm f}) = \frac{1}{\sqrt{2\pi}} \frac{\rho_{\rm b}}{M} [1 + \epsilon_{\rm bg}(\boldsymbol{q}, z)] \frac{\delta_{\rm f}(z, z_{\rm f}) - \epsilon_{\rm bg}(\boldsymbol{q}, z)}{[\sigma_M^2(z) - \sigma_0^2(z)]^{3/2}} \exp\left\{ -\frac{[\delta_{\rm f}(z, z_{\rm f}) - \epsilon_{\rm bg}(\boldsymbol{q}, z)]^2}{2[\sigma_M^2(z) - \sigma_0^2(z)]} \right\} \left| \frac{{\rm d}\sigma_M^2(z)}{{\rm d}M} \right|, \tag{23}$$

which, having averaged over the fine-grained mass fluctuations, represents a sort of coarse-grained halo counting field. Notice that the fine-grained ensemble average has replaced the original step-function operator of equation (21) by a smoother function, which can then be consistently expanded in series of the background field, as shown below.

Let us stress that the expression in equation (23) is just the Eulerian analogue of equation (10) in MW, but the field $\epsilon_{\rm bg}$ is here a true random field, and so is the process $N_{\rm h}^{\rm E}$. The knowledge of $N_{\rm h}^{\rm E}$ allows us to define the Eulerian halo number density fluctuation as

$$\delta_{\rm h}^{\rm E}(\boldsymbol{q}, z|M, z_{\rm f}) \equiv \frac{N_{\rm h}^{\rm E}(\boldsymbol{q}, z|M, z_{\rm f}) - \langle N_{\rm h}^{\rm E}(\boldsymbol{q}, z|M, z_{\rm f}) \rangle_{\epsilon_{\rm bg}}}{\langle N_{\rm h}^{\rm E}(\boldsymbol{q}, z|M, z_{\rm f}) \rangle_{\epsilon_{\rm bg}}} \equiv {\rm b}^{\rm E}(\boldsymbol{q}, z|M, z_{\rm f})\, \epsilon_{\rm bg}(\boldsymbol{q}, z), \tag{24}$$

where we introduced the Eulerian 'bias field' ${\rm b}^{\rm E}(\boldsymbol{q}, z|M, z_{\rm f})$. The second equality in the above equation does not mean that the Eulerian fluctuation field $\delta_{\rm h}^{\rm E}$ is proportional to the background density field $\epsilon_{\rm bg}$. In fact, $\rm b^E$ in general depends upon $\epsilon_{\rm bg}$ itself. Its functional dependence can be understood by expanding $N_{\rm h}^{\rm E}(\boldsymbol{q}, z|M, z_{\rm f})$ in powers of $\epsilon_{\rm bg}$ to obtain

$$\begin{aligned} \delta_{\rm h}^{\rm E}(\boldsymbol{q}, z|M, z_{\rm f}) &= b_1^{\rm E}(z|M, z_{\rm f})\, \epsilon_{\rm bg}(\boldsymbol{q}, z) + \frac{1}{2} b_2^{\rm E}(z|M, z_{\rm f})\, \epsilon_{\rm bg}^2(\boldsymbol{q}, z) + \dots \\ &= [1 + b_1^{\rm L}(z|M, z_{\rm f})]\, \epsilon_{\rm bg}(\boldsymbol{q}, z) + \frac{1}{2} [b_2^{\rm L}(z|M, z_{\rm f}) + 2 b_1^{\rm L}(z|M, z_{\rm f})]\, \epsilon_{\rm bg}^2(\boldsymbol{q}, z) + \dots, \end{aligned} \tag{25}$$

where, for $\sigma_M^2 \gg \sigma_0^2$, the first- and second-order Lagrangian bias factors $b_1^{\rm L}$ and $b_2^{\rm L}$ are those of equations (16) and (17) respectively. Accounting for the transformation from the Lagrangian to the Eulerian distribution (e.g. Kofman et al. 1992), one has $\langle N_{\rm h}^{\rm E}(\boldsymbol{q}, z|M, z_{\rm f}) \rangle_{\epsilon_{\rm bg}} = n_{\rm PS}(z|M, z_{\rm f})$. It can be useful to give explicit expressions for the first two Eulerian bias parameters of linear theory:

$$b_1^{\rm E}(z|M, z_{\rm f}) = 1 + \frac{D(z_{\rm f})}{D(z)} \left[\frac{\delta_{\rm c}}{D(z_{\rm f})^2 \sigma_M^2} - \frac{1}{\delta_{\rm c}} \right], \tag{26}$$

$$b_2^{\rm E}(z|M, z_{\rm f}) = \frac{1}{D(z)^2 \sigma_M^2} \left[\frac{\delta_{\rm c}^2}{D(z_{\rm f})^2 \sigma_M^2} - 3 \right] + \frac{2D(z_{\rm f})}{D(z)} \left[\frac{\delta_{\rm c}}{D(z_{\rm f})^2 \sigma_M^2} - \frac{1}{\delta_{\rm c}} \right]. \tag{27}$$

The set of linear theory Eulerian bias factors $b_\ell^{\rm E}(z)$ can be obtained from the Lagrangian ones according to the general rule

$$b_\ell^{\rm E} = \ell\, b_{\ell-1}^{\rm L} + b_\ell^{\rm L}, \tag{28}$$

with $b_{\ell=0}^{\rm L} \equiv 1$.

The same method can be applied to the Lagrangian expression, in the sense that we can obtain similarly,

$$N_{\rm h}^{\rm L}(\boldsymbol{q}, z|M, z_{\rm f}) = \frac{1}{\sqrt{2\pi}} \frac{\rho_{\rm b}}{M} \frac{\delta_{\rm f}(z, z_{\rm f}) - \epsilon_{\rm bg}(\boldsymbol{q}, z)}{[\sigma_M^2(z) - \sigma_0^2(z)]^{3/2}} \exp\left\{ -\frac{[\delta_{\rm f}(z, z_{\rm f}) - \epsilon_{\rm bg}(\boldsymbol{q}, z)]^2}{2[\sigma_M^2(z) - \sigma_0^2(z)]} \right\} \left| \frac{{\rm d}\sigma_M^2(z)}{{\rm d}M} \right|. \tag{29}$$

One has, exactly, $\langle N_{\rm h}^{\rm L}(\boldsymbol{q}, z|M, z_{\rm f})\rangle_{\epsilon_{\rm bg}} = \langle \mathcal{N}_{\rm h}^{\rm L}(\boldsymbol{q}, z|M, z_{\rm f})\rangle = n_{\rm PS}(z|M, z_{\rm f})$. By expanding the coarse-grained Lagrangian counting field $N_{\rm h}^{\rm L}(\boldsymbol{q}, z|M, z_{\rm f})$ we can define Lagrangian bias factors at any order. For $\sigma_M^2 \gg \sigma_0^2$ these turn out to be identical to those deriving from the expansion of the halo correlation in Lagrangian space, equation (14). This suggests, however, that these bias factors can be used to describe halo clustering on distances $r > R$, without any further restriction introduced by the background scale R_0.

The very fact that, for practical purposes, we can replace the exact operator $\mathcal{N}_{\rm h}^{\rm E}$ by the locally averaged one $N_{\rm h}^{\rm E}$ demonstrates that the MW treatment can be made self-consistent, provided that their small-scale density field is replaced by the peak field, and that the value of the threshold is modified accordingly. Most importantly, our local averaging procedure implies that, up to the scale R_0, we are indeed correctly accounting for the cloud-in-cloud problem. This is because at each point $\boldsymbol{q}$, characterized by a random value of the background field $\epsilon_{\rm bg}(\boldsymbol{q})$, the coarse-grained stochastic process $N_{\rm h}^{\rm E}(\boldsymbol{q}, z|M, z_{\rm f})$ (and its Lagrangian equivalent) actually represents the local mean mass function, for which the cloud-in-cloud problem is exactly solved in terms of first passage 'times' across the local barrier $\delta_{\rm f}(z, z_{\rm f}) - \epsilon_{\rm bg}(\boldsymbol{q}, z)$, with initial condition $\epsilon_{\rm pk}(\boldsymbol{q}, z) = 0$ at $R = R_0$. Therefore, with the aim of calculating correlations on lags $r \gg R_0$, we can safely state that our coarse-grained halo counting fields are unaffected by the cloud-in-cloud problem.

The shift by 1 of the linear bias factor, here implied by the transformation from the Lagrangian to the Eulerian world, was also noticed in the weighted bias approach by Catelan et al. (1994, their equation 21), where an underlying lognormal distribution was assumed to avoid negative-mass events.

The above expression for $b_1^{\rm E}(z|M, z_{\rm f})$ coincides with the formula by MW (their equation 20), who, however, only presented it for $z = 0$. As noticed by MW, an important feature of this linear bias is that it predicts that large-mass objects (actually those characterized by $\sigma_M < t_{\rm f}$) are biased with respect to the mass ($b_1^{\rm E} > 1$), while small-mass ones ($\sigma_M > t_{\rm f}$) are antibiased ($b_1^{\rm E} < 1$). Haloes with mass close to the characteristic one, M_*, have non-vanishing linear bias, unlike the Lagrangian case. As we will see in Section 3.1, this one-to-one classification of biased and antibiased objects according to their mass is no longer valid in the non-linear regime, as the shear field at the Lagrangian location of the halo also contributes to the determination of its Eulerian bias factor.

The effect of merging can be easily accommodated into this scheme. In the real Universe, haloes undergo merging at some finite rate, which can be suitably modelled (e.g. Lacey & Cole 1993). As mentioned above, in the simple PS theory such a rate is actually infinite, for infinite mass resolution, implying that only haloes 'just formed' can survive, so that $z_{\rm f} = z$. So, if one gives up singling out the individuality of haloes selected at different thresholds, i.e., with different formation redshifts $z_{\rm f} \geq z$, one immediately obtains (e.g. Matarrese et al. 1997)

$$b_1^{\rm E}(z|M) = 1 + \left[\frac{\delta_{\rm c}}{D(z)^2\sigma_M^2} - \frac{1}{\delta_{\rm c}}\right], \tag{30}$$

which implies a quadratic redshift dependence in the Einstein–de Sitter universe,

$$b_1^{\rm E}(z|M) = 1 + \left[\frac{\delta_{\rm c}(1+z)^2}{\sigma_M^2} - \frac{1}{\delta_{\rm c}}\right]. \tag{31}$$

The latter form coincides with the result by Cole & Kaiser (1989, their equation 6), who, however, define the bias factor of haloes at redshift z with respect to the mass fluctuation at the present time, which then scales the latter expression by a factor $(1+z)^{-1}$.

On the other hand, for fixed $z_{\rm f}$ and varying z, i.e., for objects which survived till the epoch z after their birth at $z_{\rm f}$, the Eulerian bias of equation (26) gets a completely different evolution, namely

$$b_1^{\rm E}(z|M) = 1 + \frac{D(z_{\rm f})}{D(z)}\,[b_1^{\rm E}(z_{\rm f}|M) - 1], \tag{32}$$

which implies a linear redshift dependence in the Einstein–de Sitter case,

$$b_1^{\rm E}(z|M) = 1 + \frac{1+z}{1+z_{\rm f}}\,[b_1^{\rm E}(z_{\rm f}|M) - 1]. \tag{33}$$

The latter form coincides with that obtained by Dekel (1986), Dekel & Rees (1987), Nusser & Davis (1994) and Fry (1996). This relation can be relevant for galaxies which were conserved in number after their formation, i.e., that maintained their individuality even after their hosting haloes merged.

It is trivial, at this point, to obtain the Eulerian halo–halo correlation function within our approximations. For lags $r \gg R_0$, one has

$$\xi^{\rm E}_{\rm hh}(r,\,z|M_1,\,z_1;\,M_2,\,z_2)=b^{\rm E}_1(z|M_1,\,z_1)\,b^{\rm E}_1(z|M_2,\,z_2)\,\xi_{\rm m}(r,\,z)+\frac{1}{2}\,b^{\rm E}_2(z|M_1,\,z_1)\,b^{\rm E}_2(z|M_2,\,z_2)\,\xi^2_{\rm m}(r,\,z)+\ldots \tag{34}$$

The main limitation of this formula, however, is that it only provides a link between the Eulerian halo correlation function and that of the mass within linear theory. What one would really need, instead, is a similar relation in the fully non-linear regime. This problem will be solved in the next section.

3 HALO COUNTING AND NON-LINEAR DYNAMICS: EULERIAN DESCRIPTION

One can derive a general expression for the Eulerian halo-to-mass bias by integrating the continuity equations for the mass and for the halo number density, assuming that haloes move according to the velocity field determined by the matter. The Lagrangian analysis carried out in the previous section is crucial to the present purposes, since it allows for the natural initial conditions necessary to integrate the Eulerian equations. As we will show below, the Eulerian halo-to-mass bias obtained in such a way holds for any cosmology and in any dynamical regime. This turns out to be a remarkable generalization of the biasing proposed by Cole & Kaiser (1989) and MW.

3.1 Eulerian bias from dynamical fluid equations

Let us consider the mass density fluctuation field $\delta[\boldsymbol{x},\,\tau(z)]=\delta(\boldsymbol{x},\,z)$ which obeys the mass conservation equation

$$\frac{{\rm d}\delta}{{\rm d}\tau}=-(1+\delta)\,\nabla\cdot\boldsymbol{v}, \tag{35}$$

where τ is the conformal time of the background cosmology, and the differential operator ${\rm d}/{\rm d}\tau\equiv\partial/\partial\tau+\boldsymbol{v}\cdot\nabla$ is the convective derivative. The peculiar velocity field $\boldsymbol{v}\equiv{\rm d}\boldsymbol{x}/{\rm d}\tau$ satisfies the Euler equation ${\rm d}\boldsymbol{v}/{\rm d}\tau+(a'/a)\,\boldsymbol{v}=-\nabla\phi_{\rm g}$, where a is the expansion factor, and a prime denotes differentiation with respect to τ. For later convenience, let us also define the scaled peculiar velocity $\boldsymbol{u}\equiv{\rm d}\boldsymbol{x}/{\rm d}D=\boldsymbol{v}/D'$. The peculiar gravitational potential $\phi_{\rm g}$ is determined by the matter distribution via the cosmological Poisson equation $\nabla^2\phi_{\rm g}=4\pi G a^2\rho_{\rm b}(\tau)\,\delta$, where $\rho_{\rm b}(\tau)$ is the background mean density at time τ. If we assume that our halo population of mass M and formation redshift $z_{\rm f}$ is conserved in time, and evolves exclusively under the influence of gravity, its number density fluctuation $\delta_{\rm h}(\boldsymbol{x},\,z)=\delta_{\rm h}(\boldsymbol{x},\,z|M,\,z_{\rm f})$ has to satisfy the continuity equation (e.g. Fry 1996)

$$\frac{{\rm d}\delta_{\rm h}}{{\rm d}\tau}=-(1+\delta_{\rm h})\,\nabla\cdot\boldsymbol{v}, \tag{36}$$

from which, eliminating the expansion scalar $\nabla\cdot\boldsymbol{v}$, we obtain

$$\frac{{\rm d}\,\ln(1+\delta_{\rm h})}{{\rm d}\tau}=\frac{{\rm d}\,\ln(1+\delta)}{{\rm d}\tau}. \tag{37}$$

This equation can be integrated exactly in terms of Lagrangian quantities, and the solution reads

$$1+\delta_{\rm h}(\boldsymbol{x},\,z)=[1+\delta_{\rm h}(\boldsymbol{q})]\,[1+\delta(\boldsymbol{x},\,z)] \tag{38}$$

(see also the discussion in Peacock & Dodds 1994), where $\boldsymbol{q}$ is the Lagrangian position corresponding to the Eulerian one via $\boldsymbol{x}(\boldsymbol{q},\,z)=\boldsymbol{q}+\boldsymbol{S}(\boldsymbol{q},\,z)$, with $\boldsymbol{S}(\boldsymbol{q},\,z)$ the displacement vector. In equation (38), by $\delta_{\rm h}(\boldsymbol{q})=\delta_{\rm h}(\boldsymbol{q}|M,\,z_{\rm f})$ we mean the Lagrangian halo density fluctuation, whereas, for simplicity, we assumed that $\lim_{z\to\infty}\delta[\boldsymbol{x}(\boldsymbol{q},\,z),\,z]\equiv\delta(\boldsymbol{q})=0$, i.e., that the mass was initially uniformly distributed (this amounts to taking purely growing-mode initial perturbations). Defining the Eulerian halo bias field through

$$\delta_{\rm h}(\boldsymbol{x},\,z)\equiv{\rm b}^{\rm E}(\boldsymbol{x},\,z)\,\delta(\boldsymbol{x},\,z), \tag{39}$$

we end up with the *exact* relation

$${\rm b}^{\rm E}(\boldsymbol{x},\,z)=1+\frac{1+\delta(\boldsymbol{x},\,z)}{\delta(\boldsymbol{x},\,z)}\,\delta_{\rm h}(\boldsymbol{q}). \tag{40}$$

The key problem now is how to calculate the field $\delta_{\rm h}(\boldsymbol{q})$. We cannot simply take the Lagrangian halo distribution as $\delta_{\rm h}(\boldsymbol{q})={\rm b}^{\rm L}(\boldsymbol{q})\,\delta(\boldsymbol{q})$, because $\delta(\boldsymbol{q})=0$; thus we are forced to adopt some limiting procedure. The specify the Lagrangian halo

distribution, we can take advantage of the results of Section 2. By definition, the Lagrangian distribution of nascent haloes of mass M and formation epoch z_f is given by

$$\delta_h(\boldsymbol{q}|M, z_f) \equiv \lim_{z\to\infty} \mathrm{b}^E(\boldsymbol{q}, z|M, z_f)\,\epsilon_{bg}(\boldsymbol{q}, z) \equiv \mathrm{b}_0^L(\boldsymbol{q}|M, z_f)\,\epsilon_0(\boldsymbol{q}), \tag{41}$$

where $b_0^L(\boldsymbol{q}|M, z_f)$ is the Lagrangian halo bias field. Once again, let us stress that the second equality in the latter equation does not mean at all that $\delta_h(\boldsymbol{q})$ is proportional to $\epsilon_0(\boldsymbol{q})$. In fact, b^L is in general a functional of the background density field. To understand the above equation, one has to remember that, at sufficiently early times, the expression for the Eulerian bias field obtained in linear theory becomes exact (as linear theory gets more and more accurate), and $\delta(\boldsymbol{x}, z)\to\epsilon_{bg}(\boldsymbol{x}, z) = D(z)\,\epsilon_0(\boldsymbol{q})$, as $z\to\infty$. Because of our normalization of D, here $\epsilon_0(\boldsymbol{q})$ is the mass density fluctuation linearly extrapolated to the presentt time and filtered on the background scale R_0. The background smoothing scale R_0 actually has a twofold role in our analysis. In the linear theory approach of Section 2 it was introduced and required to be much larger than the halo size, in order to get a self-consistent definition of halo counting fields, with the desirable feature of being free of the cloud-in-cloud problem. In the present non-linear analysis, however, the background mass scale must be chosen large enough to ensure that the halo velocity field coincides with the one of the matter.

The Lagrangian density contrast of haloes identified by a PS-type algorithm can be obtained from equation (29) as $\delta_h(\boldsymbol{q}|M, z_f) = N_h^L(\boldsymbol{q}, z|M, z_f)/n_{PS}(z|M, z_f) - 1$, which leads to

$$\delta_h(\boldsymbol{q}|M, z_f) = \left[1 - \frac{D(z_f)\,\epsilon_0(\boldsymbol{q})}{\delta_c}\right]\left(1 - \frac{\sigma_0^2}{\sigma_M^2}\right)^{-3/2} \exp\left[-\frac{\epsilon_0(\boldsymbol{q})^2 - 2\epsilon_0(\boldsymbol{q})\,\delta_c/D(z_f) + \delta_c^2\sigma_0^2/D(z_f)^2\sigma_M^2}{2(\sigma_M^2 - \sigma_c^2)}\right] - 1. \tag{42}$$

For $\sigma_M^2 \gg \sigma_0^2$ this expression simplifies to

$$\delta_h(\boldsymbol{q}|M, z_f) = \left[1 - \frac{D(z_f)\,\epsilon_0(\boldsymbol{q})}{\delta_c}\right] \exp\left[-\frac{\epsilon_0(\boldsymbol{q})^2 - 2\epsilon_0(\boldsymbol{q})\,\delta_c/D(z_f)}{2\sigma_M^2}\right] - 1 = \sum_{\ell=1}^{\infty} \frac{b_{0\ell}^L(M, z_f)}{\ell!}\,\epsilon_0(\boldsymbol{q})^\ell. \tag{43}$$

The first four Lagrangian bias factors evaluated at $z=0$ read

$$b_{01}^L(M, z_f) = D(z_f)\left[\frac{\delta_c}{D(z_f)^2\sigma_M^2} - \frac{1}{\delta_c}\right], \tag{44}$$

$$b_{02}^L(M, z_f) = \frac{1}{\sigma_M^2}\left[\frac{\delta_c^2}{D(z_f)^2\sigma_M^2} - 3\right], \tag{45}$$

$$b_{03}^L(M, z_f) = \frac{D(z_f)}{\sigma_M^2}\left[\frac{\delta_c^3}{D(z_f)^4\sigma_M^4} - \frac{6\delta_c}{D(z_f)^2\sigma_M^2} + \frac{3}{\delta_c}\right], \tag{46}$$

$$b_{04}^L(M, z_f) = \frac{1}{\sigma_M^4}\left[\frac{\delta_c^4}{D(z_f)^4\sigma_M^4} - \frac{10\delta_c^2}{D(z_f)^2\sigma_M^2} + 15\right]. \tag{47}$$

Note that, in full generality, $b_{0\ell}^L(M, z_f) = D(z)^\ell b_\ell^L(z|M, z_f)$. Adding the further requirement that the local fluctuation on the background scale R_0 has not collapsed yet by the time of halo formation would make our object number density semipositive definite both at the Lagrangian and Eulerian level, i.e., $\delta_h \geq -1$, at any time, provided only that $\epsilon_0 \leq t_f$.

The general expression for the Lagrangian halo density contrast of equation (42) is plotted in Fig. 2 as a function of the background density field, for different halo masses. In the high-mass case, positive mass fluctuations typically correspond to positive values of the Lagrangian halo density contrast (and vice versa), while the trend is the opposite at low masses. The transition, once again, corresponds to halo masses around M_*, in which case positive values of δ_h occur only in regions with background density close to the mean. Also shown are two approximations to the Lagrangian halo density contrast obtained by expanding equation (42) up to first and second order in the background field. Except for halo masses near M_*, where a quadratic bias is clearly needed, a linear Lagrangian bias generally provides an accurate fit to $\delta_h(\boldsymbol{q})$ within the bulk of the ϵ_0 distribution.

The Eulerian bias field finally reads

$$\mathrm{b}^E(\boldsymbol{x}, z|M, z_f) = 1 + \frac{1 + \delta(\boldsymbol{x}, z)}{\delta(\boldsymbol{x}, z)}\,\mathrm{b}_0^L(\boldsymbol{q}|M, z_f)\,\epsilon_0(\boldsymbol{q}). \tag{48}$$

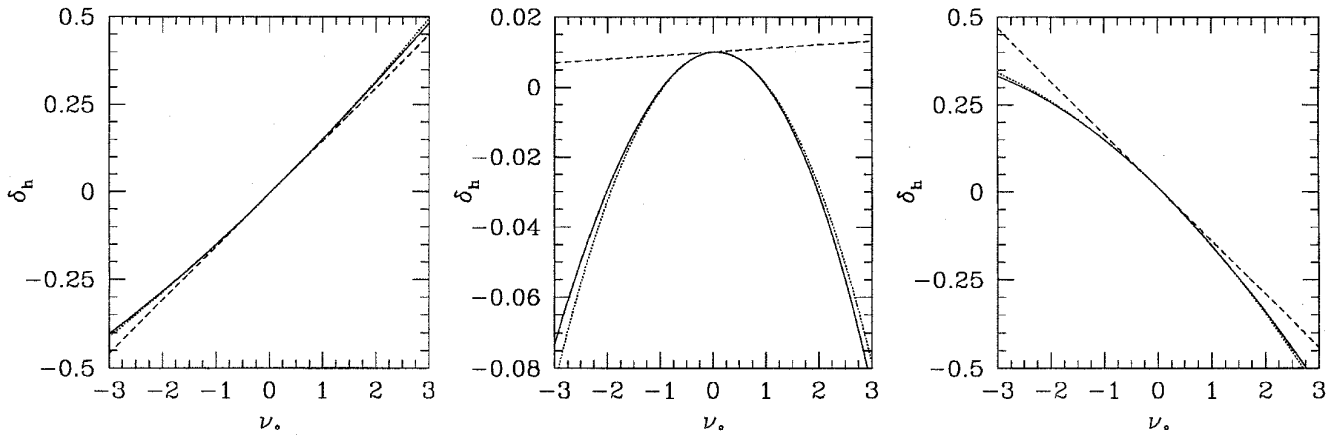

Figure 2. The exact expression for the Lagrangian halo density contrast of equation (42) (solid lines) is plotted as a function of $\nu_0 \equiv \epsilon_0/\sigma_0$. The three panels refer to values of the halo masses such that $\sigma_M^2/t_f^2 = 1/4$ (left), $\sigma_M^2/t_f^2 = 1$ (centre) and $\sigma_M^2/t_f^2 = 4$ (right), with $t_f = 1.69/D(z_f)$. The background mass scale is chosen so that $\sigma_0^2 = 0.01\,\sigma_M^2$. Also plotted are two estimates of δ_h obtained by expanding the right-hand side of equation (42) up to first (dashed lines) and second order (dotted lines) in ϵ_0. Note that, because of our choice of variables, all the curves are independent both of z_f and Ω_0.

It can be seen that, in the linear regime, where $\delta(\boldsymbol{x}, z) \approx D(z)\,\epsilon_0(\boldsymbol{q}) \ll 1$, the expression for b^E in MW (i.e., our equation 26) is recovered, provided that $b_0^L(\boldsymbol{q}|M, z_f)$ is replaced by its first-order approximation, $b_{01}^L(M, z_f)$. It is, however, important to realize that the exact expression in equation (48) implies that the Eulerian bias field of dark matter haloes $b^E(\boldsymbol{x}, z|M, z_f)$ is both non-linear, in that it depends on $\delta(\boldsymbol{x})$, and non-local, as it depends on the Lagrangian position $\boldsymbol{q}$ through $b_0^L(\boldsymbol{q}|M, z_f)\,\epsilon_0(\boldsymbol{q})$, simply because of inertia.

Our exact expression for the Eulerian halo bias (equation 48) generally involves quantitative corrections to the MW approximate bias formula. In some cases, however, the MW relation may even fail to predict the correct qualitative behaviour of the halo-to-mass bias. This is the case, in fact, of those initially underdense fluid elements in Lagrangian space, $\epsilon_0(\boldsymbol{q}) < 0$, which, after an initial expansion phase, turn around to undergo a phase of local compression, so that the corresponding Eulerian fluid element eventually becomes overdense, $\delta[\boldsymbol{x}(\boldsymbol{q}, z), z] > 0$, and collapses. This is a well-known non-linear effect caused by the shear component of the velocity field, i.e., by the tidal force of the surrounding matter. For Gaussian initial conditions, the occurrence of such an event can be estimated by the Zel'dovich approximation as affecting 42 per cent of the overall Lagrangian volume (Doroshkevich 1970; Shandarin & Zel'dovich 1984); Hui & Bertschinger (1996), using a different approximation, estimated this effect as affecting at least 39 per cent of the total Lagrangian volume. In all such cases the MW formula would incorrectly predict bias instead of antibias for halo masses $M > M_*$, and antibias instead of bias for $M < M_*$. The problem may be generally less severe than the above heuristic argument would suggest, as, at a fixed epoch z, only a smaller fraction of such Lagrangian patches have already turned around from their initial expansion; this is even more true for the large-mass haloes, which probe the underlying mass distrtibution in a more linear regime, where the MW formula gets closer to the exact one. As a tentative conclusion, let us say that one should be careful in applying the linear MW bias (i.e., our equation 26) at the Eulerian level, especially in connection with halo masses much smaller than M_*.

The most important application of equation (48) is that it allows us to generate Eulerian maps of the local comoving halo number density per unit mass, $n_{PS}(M, z_f)\,[1 + \delta_h(\boldsymbol{x}, z|M, z_f)]$, given the non-linearly evolved mass density contrast $\delta(\boldsymbol{x}, z)$ (with Lagrangian resolution R_0) and the corresponding Lagrangian mass and halo density fluctuation fields, $\epsilon_0(\boldsymbol{q})$ and $\delta_h(\boldsymbol{q}|M, z_f)$ respectively.

In order to account for halo merging, at this level, one just has to assume a suitable link between the formation and observation epochs, which, in the simple PS theory amounts to the replacement $z_f \to z$, in the above expression for b_0^L.

Recalling that mass conservation can be recast in terms of the Jacobian determinant $J \equiv \|\partial \boldsymbol{x}/\partial \boldsymbol{q}\|$ of the mapping $\boldsymbol{q} \to \boldsymbol{x}$, as $1 + \delta[\boldsymbol{x}(\boldsymbol{q}, z), z] = J(\boldsymbol{q}, z)^{-1}$, one finds the exact relation

$$b^E[\boldsymbol{x}(\boldsymbol{q}, z), z|M, z_f] = 1 + [1 - J(\boldsymbol{q}, z)]^{-1}\, b_0^L(\boldsymbol{q}|M, z_f)\,\epsilon_0(\boldsymbol{q}). \tag{49}$$

It can be useful to illustrate the meaning of this expression by considering various approximations to the evolution of the mass density in the non-linear regime, i.e., to the particle trajectories $\boldsymbol{x}(\boldsymbol{q}, z)$. Such approximation schemes should be thought of, not as self-consistent perturbative approaches to the actual dynamics, but as 'clever tricks' able to catch some aspects of the true dynamics, at least in the mildly non-linear regime. A detailed and systematic comparison of the performance of several approximations for different choices of the initial conditions has been made by Sathyaprakash et al. (1995).

3.1.1 *Zel'dovich approximation*

In the Zel'dovich approximation (ZEL) (Zel'dovich 1970) the displacement vector is $\boldsymbol{S}=-D\nabla_{\boldsymbol{q}}\varphi_0(\boldsymbol{q})$, where $\varphi_0(\boldsymbol{q})$ is the linear peculiar gravitational potential, suitably rescaled so that $\nabla^2_{\boldsymbol{q}}\varphi_0(\boldsymbol{q})=\epsilon_0(\boldsymbol{q})$. Indicating by $\lambda_\alpha(\boldsymbol{q})$ ($\alpha=1, 2, 3$) the eigenvalues of the deformation tensor $\partial^2\varphi_0(\boldsymbol{q})/\partial q_\alpha\partial q_\beta$, we obtain for the Eulerian bias field

$$\mathrm{b}^{\mathrm{E}}_{\mathrm{ZEL}}[\boldsymbol{x}(\boldsymbol{q}, z), z|M, z_{\mathrm{f}}]=1+\frac{\epsilon_0(\boldsymbol{q})\,\mathrm{b}^{\mathrm{L}}_0(\boldsymbol{q}|M, z_{\mathrm{f}})}{1-\Pi^3_{\alpha=1}[1-D(z)\,\lambda_\alpha(\boldsymbol{q})]}$$

$$=1+\frac{\mathrm{b}^{\mathrm{L}}_0(\boldsymbol{q}|M, z_{\mathrm{f}})}{D(z)}\left[1-D(z)\,\frac{\mu_2(\boldsymbol{q})}{\mu_1(\boldsymbol{q})}+D(z)^2\,\frac{\mu_3(\boldsymbol{q})}{\mu_1(\boldsymbol{q})}\right]^{-1}. \tag{50}$$

Here $\mu_1(\boldsymbol{q})\equiv\lambda_1+\lambda_2+\lambda_3=\epsilon_0(\boldsymbol{q})$, $\mu_2\equiv\lambda_1\lambda_2+\lambda_1\lambda_3+\lambda_2\lambda_3$ and $\mu_3\equiv\lambda_1\lambda_2\lambda_3$ are the three invariants of the deformation tensor. If one makes the further approximation of replacing the Lagrangian bias by its first-order estimate of equation (44), it can be checked that the expression of $\mathrm{b}^{\mathrm{E}}_{\mathrm{ZEL}}$ coincides with the MW result, both at sufficiently early times ($D\ll 1$) and in the case of one-dimensional perturbations, for which $\mu_2=0=\mu_3$ and the Zel'dovich approximation represents the exact solution to the non-linear dynamics.

It is important to stress that we are not forced to take the above result as a perturbative expression. An accurate approximation to the Eulerian bias field would, in fact, consist in evolving the mass according to the truncated (on the scale M_0) Zel'dovich approximation (Kofman 1991; Kofman et al. 1992; Coles, Melott & Shandarin 1993) and using the full expression for the Lagrangian bias. Being a random field, the Eulerian halo bias is completely characterized by a probability density functional; thus for a given mass M and formation redshift z_{f} there exists a whole distribution of possible values of b^{E}, related to the particular environment where the object forms as well as to the initial conditions leading to that site. Starting from the ZEL expression in equation (50), one could explicitly obtain the probability distribution function $p(\mathrm{b}^{\mathrm{E}}_{\mathrm{ZEL}})\,\mathrm{db}^{\mathrm{E}}_{\mathrm{ZEL}}$ by integrating over the joint distribution of the invariants μ_α (an expression for the latter is given in Kofman et al. 1994). These specific applications of our results will be discussed elsewhere.

Equation (50) has the merit of clearly displaying the intrinsic non-locality of the Eulerian bias. Only in some simplified cases does there exist a local mapping between b^{E} and δ, so that an expansion of the halo density contrast in a hierarchy of Eulerian bias factors, b^{E}_1, b^{E}_2, etc., makes sense. One example is provided by the linear-theory approach of Section 2.6; further examples are given below.

3.1.2 *Frozen-flow approximation*

According to the *frozen-flow* approximation (FFA) (Matarrese et al. 1992) the Eulerian density field can be written as

$$1+\delta[\boldsymbol{x}(\boldsymbol{q}, z), z]=\exp\int_0^{D(z)}\mathrm{d}\tilde{D}\,\epsilon_0[\boldsymbol{x}(\boldsymbol{q}, \tilde{D})], \tag{51}$$

where the integral is calculated along the trajectory of the fluid element. Note that, since in the FFA shell-crossing never occurs, the mapping $\boldsymbol{q}\rightarrow\boldsymbol{x}$ can be inverted at any time. The solution (51) might be replaced in equation (48) to obtain a non-local expression for the FFA bias parameter. However, we can make a further step by noting that, for Lagrangian points $\boldsymbol{q}_*$ corresponding to local extrema of the initial gravitational potential $\nabla\boldsymbol{q}\,\varphi_0(\boldsymbol{q}_*)=\boldsymbol{0}$, FFA predicts $\boldsymbol{x}_*=\boldsymbol{x}(\boldsymbol{q}_*, z)=\boldsymbol{q}_*$, and

$$1+\delta(\boldsymbol{x}_*, z)=\exp[D(z)\,\epsilon_0(\boldsymbol{x}_*)]. \tag{52}$$

One can speculate that such points represent the preferential sites for the formation of massive haloes, which could be associated to clusters of galaxies, and use this approximate expression to obtain

$$\mathrm{b}^{\mathrm{E}}_{\mathrm{FFA}}(\boldsymbol{x}_*, z|M, z_{\mathrm{f}})\approx 1+\frac{1+\delta(\boldsymbol{x}_*, z)}{\delta(\boldsymbol{x}_*, z)}\ln[1+\delta(\boldsymbol{x}_*, z)]\,\frac{\mathrm{b}^{\mathrm{L}}_0(\boldsymbol{x}_*|M, z_{\mathrm{f}})}{D(z)}. \tag{53}$$

Expanding this expression in powers of δ, to first order we recover the MW expression, equation (26), while to second order we obtain

$$b^{\mathrm{E}}_{2\mathrm{FFA}}(z|M, z_{\mathrm{f}})=\frac{1}{D(z)^2\sigma_M^2}\left[\frac{\delta_{\mathrm{c}}^2}{D(z_{\mathrm{f}})^2\sigma_M^2}-3\right]+\frac{D(z_{\mathrm{f}})}{D(z)}\left[\frac{\delta_{\mathrm{c}}}{D(z_{\mathrm{f}})^2\sigma_M^2}-\frac{1}{\delta_{\mathrm{c}}}\right], \tag{54}$$

which differs from the linear-theory prediction of equation (27). Analogous results could be obtained using the *frozen-potential* approximation (Brainerd, Scherrer & Villumsen 1993; Bagla & Padmanabhan 1994), with the main difference that the δ

evolution would be slowed down compared to FFA. Quite interesting is the fact that the *lognormal model* by Coles & Jones (1991) assumes that the quantity $1+33.\ \delta(\boldsymbol{x}, z)$ can be always approximated by the exponential of the linear density field at the same Eulerian position, so that the expressions above for the Eulerian bias factor in FFA would apply to all Eulerian points $\boldsymbol{x}$. Of course, the validity of these approximate expressions for the bias should be checked against the results of N-body simulations.

Another way to get a local mapping between the evolved halo density field and the underlying matter perturbations is to approximate the non-linear evolution of the mass by the spherical top-hat model. This method has been followed by MW and Mo et al. (1996).

3.2 Perturbative evaluation of the Eulerian halo density contrast

In the previous section we demonstrated that the Eulerian bias is a non-linear and non-local function of the density fluctuation field. The 'non-locality', in particular, comes from the fact that the halo number density fluctuation in $\boldsymbol{x}$ is determined by the initial halo number fluctuation at the Lagrangian position $\boldsymbol{q}$, which, in turn, is related to the linear mass fluctuation in the same point, through a hierarchy of Lagrangian bias parameters. Here we want to derive an approximate expression for $\delta_{\rm h}[\boldsymbol{x}(\boldsymbol{q}, D), D]$, by applying the second-order Eulerian perturbation theory. Whenever it will be necessary to go from the Lagrangian position $\boldsymbol{q}$ to the Eulerian one, the Zel'dovich approximation will be sufficient.

Within the linear regime, the Eulerian solution of the continuity equation is simply $\delta^{(1)}(\boldsymbol{x}, D)=D\epsilon_0(\boldsymbol{x})$. The mildly non-linear regime may be approximately described by the second-order solution (Bouchet et al. 1992; Bernardeau 1994; Catelan et al. 1995)

$$\delta^{(2)}(\boldsymbol{x}, D)=\frac{1}{2}\left[1-\frac{E}{D^2}\right]\delta^{(1)}(\boldsymbol{x}, D)^2-D\boldsymbol{u}^{(1)}(\boldsymbol{x})\cdot\nabla\delta^{(1)}(\boldsymbol{x}, D)+\frac{1}{2}D^2\left[1+\frac{E}{D^2}\right]\partial_\alpha u_\beta^{(1)}(\boldsymbol{x})\,\partial_\alpha u_\beta^{(1)}(\boldsymbol{x}), \tag{55}$$

in such a way that $\delta=\delta^{(1)}+\delta^{(2)}$ and higher order corrections are neglected. Here $\boldsymbol{u}^{(1)}(\boldsymbol{x})=-\nabla\varphi_0(\boldsymbol{x})$ is the (scaled) linear peculiar velocity, and φ_0 is the (scaled) peculiar gravitational potential, linearly extrapolated to the present time. The second-order growth factor $E=E(D)$ is quite a complicated function of $D(\Omega)$ (see Appendix A for its explicit expression), but in the vicinity of $\Omega=1$ (actually in the range $0.05\leq\Omega\leq3$) it can be approximated by the expression $E\approx-\frac{3}{7}\Omega^{-2/63}D^2+\mathcal{O}[(\Omega-1)^2]$ (see Bouchet et al. 1992). Therefore the previous second-order solution is well approximated by the expression which holds in the Einstein–de Sitter universe, namely (Fry 1984)

$$\delta^{(2)}(\boldsymbol{x}, D)=\frac{5}{7}\delta^{(1)}(\boldsymbol{x}, D)^2-D\boldsymbol{u}^{(1)}(\boldsymbol{x})\cdot\nabla\delta^{(1)}(\boldsymbol{x}, D)+\frac{2}{7}D^2\partial_\alpha u_\beta^{(1)}(\boldsymbol{x})\,\partial_\alpha u_\beta^{(1)}(\boldsymbol{x}). \tag{56}$$

We want now to compute the corresponding second-order perturbative correction, $\delta_{\rm h}^{(2)}(\boldsymbol{x}, D)$, to the linear halo density fluctuation field, $\delta_{\rm h}^{(1)}(\boldsymbol{x}, D)$. From equation (38) we obtain

$$\delta_{\rm h}\approx\delta_x^{(1)}+\delta_x^{(2)}+b_1^{\rm L}\delta_q^{(1)}+b_1^{\rm L}\delta_x^{(1)}\delta_q^{(1)}+\frac{1}{2}b_2^{\rm L}\delta_q^{(1)2}, \tag{57}$$

where to maintain compact notation we write, e.g., $\delta_x^{(j)}\equiv\delta^{(j)}(\boldsymbol{x}, D)$. The Lagrangian bias factors $b_1^{\rm L}=b_1^{\rm L}(z|M, z_{\rm f})$ and $b_2^{\rm L}=b_2^{\rm L}(z|M, z_{\rm f})$ are those given in equations (16) and (17). Notice that the perturbative expansion of δ_x holds at sufficiently early times and/or large scales, while the validity of the expansion of $\delta_{\rm h}(\boldsymbol{q})$ in powers of $\epsilon_0(\boldsymbol{q})$ is based on assuming a suitably large smoothing radius R_0 on the background field $\epsilon(\boldsymbol{q})$.

The key point is that the first-order density field at the Lagrangian position $\boldsymbol{q}$ originates a non-local term, when written at the corresponding Eulerian position $\boldsymbol{x}$. Using the Zel'dovich approximation $\boldsymbol{x}=\boldsymbol{q}+D\boldsymbol{u}^{(1)}$, one obtains $\delta_q^{(1)}=\delta_x^{(1)}-D\boldsymbol{u}^{(1)}\cdot\nabla\delta_x^{(1)}$. Finally, defining $\delta_{\rm h}=\delta_{\rm h}^{(1)}+\delta_{\rm h}^{(2)}$, one gets $\delta_{\rm h}^{(1)}=(1+b_1^{\rm L})\,\delta^{(1)}$ and

$$\delta_{\rm h}^{(2)}=\left[\frac{1}{2}\left(1-\frac{E}{D^2}\right)+b_1^{\rm L}+\frac{1}{2}b_2^{\rm L}\right]\delta^{(1)2}-D(1+b_1^{\rm L})\,\boldsymbol{u}^{(1)}\cdot\nabla\delta^{(1)}+\frac{1}{2}D^2\left(1+\frac{E}{D^2}\right)\partial_\alpha u_\beta^{(1)}\,\partial_\alpha u_\beta^{(1)}. \tag{58}$$

Thus the non-locality has the effect of modifying the inertia term $\boldsymbol{u}^{(1)}\cdot\nabla\delta^{(1)}$, which gets multiplied by the factor $(1+b_1^{\rm L})$. The dynamical properties of the random field $\delta_{\rm h}$ may be equivalently analysed in terms of its Fourier transform $\tilde{\delta}_{\rm h}(\boldsymbol{k}, t)$ where $\boldsymbol{k}$ is the comoving wavevector. Thus the second-order solution (58) may be written as

$$\tilde{\delta}_{\rm h}^{(2)}(\boldsymbol{k}, D)=\int\frac{{\rm d}\boldsymbol{k}_1{\rm d}\boldsymbol{k}_2}{(2\pi)^3}\,\delta_D(\boldsymbol{k}_1+\boldsymbol{k}_2-\boldsymbol{k})\,\mathscr{K}_{\rm S}^{(2)}(\boldsymbol{k}_1, \boldsymbol{k}_2; b_1^{\rm L}, b_2^{\rm L}; \Omega)\,\tilde{\delta}_1(\boldsymbol{k}_1, D)\,\tilde{\delta}_1(\boldsymbol{k}_2, D), \tag{59}$$

where the symmetrized kernel $\mathscr{H}_{\rm S}^{(2)}$ reads

$$\mathscr{H}_{\rm S}^{(2)}(\boldsymbol{k}_1, \boldsymbol{k}_2; b_1^{\rm L}, b_2^{\rm L}, \Omega) \equiv \left[\frac{1}{2}\left(1-\frac{E}{D^2}\right)+b_1^{\rm L}+\frac{1}{2}\,b_2^{\rm L}\right]+\frac{1+b_1^{\rm L}}{2}\left(\frac{k_1}{k_2}+\frac{k_2}{k_1}\right)\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_1k_2}+\frac{1}{2}\left(1+\frac{E}{D^2}\right)\left(\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_1k_2}\right)^2. \tag{60}$$

The corresponding kernel for the Einstein–de Sitter case reads

$$\mathscr{H}_{\rm S}^{(2)}(\boldsymbol{k}_1, \boldsymbol{k}_2; b_1^{\rm L}, b_2^{\rm L}, \Omega=1) \equiv \left[\frac{5}{7}+b_1^{\rm L}+\frac{1}{2}\,b_2^{\rm L}\right]+\frac{1+b_1^{\rm L}}{2}\left(\frac{k_1}{k_2}+\frac{k_2}{k_1}\right)\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_1k_2}+\frac{2}{7}\left(\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_1k_2}\right)^2. \tag{61}$$

3.3 Halo bispectrum and skewness

A possible application of these results is the evaluation of the bispectrum and corresponding skewness of the halo distribution. A related calculation has been performed by Fry (1996), who assumed the bias to be local in Eulerian space at $z_{\rm f}$. It should be clear that our model is quite different from the local Eulerian bias prescription applied to the analysis of the skewness by Fry & Gaztañaga (1993). Moreover, the latter treatment, unlike ours, lacks of any prediction for the value of the different bias parameters. We recall that the value of the gravitationally induced skewness of the mass is

$$S=\frac{\langle\delta^3\rangle}{\langle\delta^2\rangle^2}=4-2\,\frac{E}{D^2}, \tag{62}$$

for unfiltered fields, and

$$S(\mathscr{R})=\frac{\langle\delta^3_{\mathscr{R}}\rangle}{\langle\delta^2_{\mathscr{R}}\rangle^2}=4-2\,\frac{E}{D^2}-\gamma(\mathscr{R}), \tag{63}$$

for a spherical top-hat filter, where $\gamma\equiv-{\rm d}\ln\sigma(\mathscr{R})^2/{\rm d}\ln\mathscr{R}$ (Bernardeau 1994). The smoothing radius $\mathscr{R}$ should not be confused with R, defining the halo mass: one is obviously interested in computing the skewness on a smoothing scale much larger than the typical size of the single objects. In the Einstein–de Sitter universe, and for a scale-free power spectrum with spectral index n, the latter reduces to $S(\mathscr{R})=34/7-(n+3)$, for $-3\le n\le 1$.

The derivation of the halo skewness $\langle\delta_{\rm h}^3\rangle\approx 3\langle\delta_{\rm h}^{(1)2}\delta_{\rm h}^{(2)}\rangle$ is simple. Assuming that the Eulerian halo density field is smoothed by a top-hat filter, the halo skewness parameter $S_{\rm h}$ is, for a generic value of Ω,

$$S_{\rm h}(\mathscr{R}; z, \Omega)=3\,\frac{\langle\delta_{\rm h}^{(1)2}\delta_{\rm h}^{(2)}\rangle}{\langle\delta_{\rm h}^{(1)2}\rangle^2}=\frac{4-2\dfrac{E}{D^2}+6b_1^{\rm L}(z|M, z_{\rm f})+3b_2^{\rm L}(z|M, z_{\rm f})-[1+b_1^{\rm L}(z|M, z_{\rm f})]\,\gamma(\mathscr{R})}{[1+b_1^{\rm L}(z|M, z_{\rm f})]^2}. \tag{64}$$

It is of interest to write the halo skewness in the Einstein–de Sitter universe and for a scale-free linear power spectrum,

$$S_{\rm h}(n; z, \Omega=1)=\frac{\frac{34}{7}+6b_1^{\rm L}(z|M, z_{\rm f})+3b_2^{\rm L}(z|M, z_{\rm f})-(n+3)\,[1+b_1^{\rm L}(z|M, z_{\rm f})]}{[1+b_1^{\rm L}(z|M, z_{\rm f})]^2}. \tag{65}$$

As for the mass skewness, the dependence on the smoothing scale $\mathscr{R}$ now simply translates into a dependence on the spectral index n. The asymptotic value of $S_{\rm h}(n; z, \Omega=1)$ for a fixed formation redshift $z_{\rm f}$, is $34/7-(n+3)$ as $z\to-1$. This limit gives the value of the underlying mass skewness: in the absence of merging the haloes would evolve towards an unbiased distribution in the far future. The skewness parameter is shown in Fig. 3 for different values of Ω_0 and for a scale-free model with $n=-2$. For objects observed at the present time, $z=0$, we vary the collapse epoch $z_{\rm f}$, which may simulate different models of galaxy formation inside dark haloes. By varying together $z=z_{\rm f}$, we instead show the skewness evolution in the instantaneous merging model. We also consider the case of varying only z: this gives the evolution of the skewness in a model in which the objects did not suffer any merging after their formation at $z_{\rm f}$. Finally, we show the evolution of the skewness parameter of filtered mass fluctuations; note that the Einstein–de Sitter case displays no redshift dependence, simply because of self-similarity; for sensible values of $\Omega_0\neq 1$ also the mass skewness of non-flat Friedmann models experiences very little evolution. The redshift dependence of $S_{\rm h}$ is therefore mostly due to that of the Lagrangian bias factors. Quite interesting, in this respect, is the fact that the halo skewness plotted in the two top panels of Fig. 3 displays a turning point in its redshift dependence: this typically occurs when $M\approx M_*(z_{\rm f})$.

Of particular interest is also the expression for the halo bispectrum $B_{\rm h}$ defined by the relation

$$\langle\tilde\delta_{\rm h}(\boldsymbol{k}_2, D)\,\tilde\delta_{\rm h}(\boldsymbol{k}_2, D)\,\tilde\delta_{\rm h}(\boldsymbol{k}_3, D)\rangle\equiv(2\pi)^3\delta_D(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3)\,B_{\rm h}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3; D). \tag{66}$$

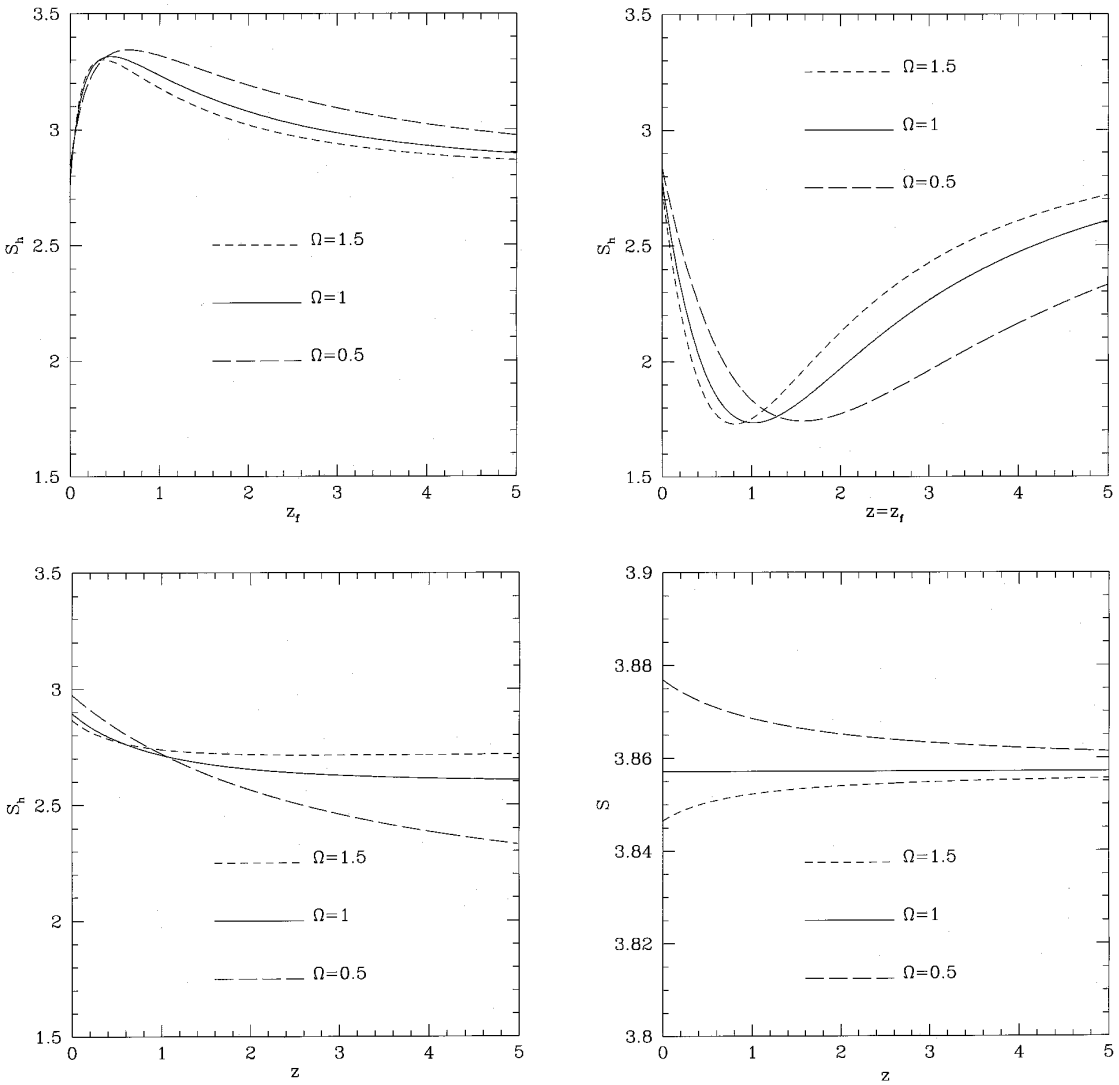

Figure 3. The filtered skewness parameter is plotted, for $\Omega_0=0.5$, 1, 1.5, for a scale-free model with $n=-2$. The halo masses are selected with the same linear mass variance $\sigma_M^2=10$, corresponding to the same present-day bias parameters. We take everywhere $\delta_c=1.69$. The top-left panels refers to objects observed at $z=0$, with varying formation redshift z_f. The top-right panels shows the effect of varying simultaneously $z=z_f$. In the bottom-left panel we fix $z_f=5$ and look at different observation redshifts $z\leq z_f$. The bottom-right panel, finally, shows the evolution of the skewness parameter of filtered mass fluctuations.

The leading term shows the characteristic hierarchical pattern

$$B_h(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3; D)=D^4[1+b_1^L(z|M, z_f)]^2[2\mathscr{H}_S^{(2)}(\boldsymbol{k}_1, \boldsymbol{k}_2; b_1^L, b_2^L, \Omega)P(k_1)P(k_2)+\text{cyclic terms}], \tag{67}$$

where $P(k)$ is the primordial density power spectrum defined by $\langle\tilde{\delta}_1(k_1, D)\tilde{\delta}_1(k_2, D)\rangle=(2\pi)^3\delta_D(\boldsymbol{k}_1+\boldsymbol{k}_2)D^2P(k_1)$, and the two cyclic terms are obtained by the substitutions $\boldsymbol{k}_1\rightarrow\boldsymbol{k}_2$, $\boldsymbol{k}_1\rightarrow\boldsymbol{k}_3$ and $\boldsymbol{k}_2\rightarrow\boldsymbol{k}_3$. Typically, as for the hierarchical mass bispectrum, the halo bispectrum is largely scale-dependent, while its dependence on the $\boldsymbol{k}$-shape is rather weak. One way to eliminate the scale dependence and look at the residual shape dependence is to analyse the 'effective' bispectrum amplitude Q (Fry 1984),

$$Q\equiv\frac{B_h(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3; D)}{P_h(k_1, D)P_h(k_2, D)+P_h(k_1, D)P_h(k_2, D)+P_h(k_2, D)P_h(k_3, D)}. \tag{68}$$

The halo power spectrum is biased with respect to the mass one, $P_h(k, D)=D^2[1+b_1^L(z)]^2P(k)$. For a power-law spectrum, the amplitude Q generally depends on the spectral index n, owing to the wavenumber modulation introduced by the kernel $\mathscr{H}_S^{(2)}(\boldsymbol{k}_1, \boldsymbol{k}_2)$ (cf. Fig. 4). For equilateral triangle configurations, Q gets an n-independent value, namely

$$Q_{\rm eq}(\Omega; z)=\frac{\frac{1}{4}\left[1-3\,\dfrac{E}{D^2}\right]+2b_1^{\rm L}(z|M,\,z_{\rm f})+b_2^{\rm L}(z|M,\,z_{\rm f})}{[1+b_1^{\rm L}(z|M,\,z_{\rm f})]^2}, \tag{69}$$

reducing to

$$Q_{\rm eq}(\Omega=1; z)=\frac{\frac{4}{7}+2b_1^{\rm L}(z|M,\,z_{\rm f})+b_2^{\rm L}(z|M,\,z_{\rm f})}{[1+b_1^{\rm L}(z|M,\,z_{\rm f})]^2}, \tag{70}$$

in the Einstein–de Sitter universe.

3.4 Local Lagrangian bias

So far, our model has been treated as being fully predictive. Once the cosmological scenario and the structure formation model have been fixed, our algorithm contains no fitting parameters. This is because we used a local version of the PS theory to generate the Lagrangian halo density contrast. One could, however, take a more general point of view and assume that the Lagrangian halo density contrast $\delta_{\rm h}(\boldsymbol{q})$ is specified in terms of the linear background density field $\epsilon_{\rm bg}(\boldsymbol{q}, z)=D(z)\,\epsilon_0(\boldsymbol{q})$ by a set of unknown bias parameters $\hat{b}_\ell^{\rm L}(z)$, as follows:

$$\delta_{\rm h}(\boldsymbol{q})=\sum_{\ell=1}^{\infty}\frac{\hat{b}_{0\ell}^{\rm L}}{\ell\,!}\,\epsilon_0(\boldsymbol{q})^\ell=\sum_{\ell=1}^{\infty}\frac{\hat{b}_\ell^{\rm L}(z)}{\ell\,!}\,\epsilon_{\rm bg}(\boldsymbol{q}, z)^\ell. \tag{71}$$

Defining now $b_1\equiv b_1(z)=1+\hat{b}_1^{\rm L}(z)$ and $b_2\equiv b_2(z)=2\hat{b}_1^{\rm L}(z)+\hat{b}_2^{\rm L}(z)$, according to equation (28), and replacing these expansions in our previous treatment, we recover the general expression (59) for the second-order halo density contrast, with the more general kernel

$$\mathscr{H}_{\rm S}^{(2)}(\boldsymbol{k}_1, \boldsymbol{k}_3; b_1, b_2, \Omega)=\frac{1}{2}\left[\left(1-\frac{E}{D^2}\right)+b_2\right]+\frac{b_1}{2}\left(\frac{k_1}{k_2}+\frac{k_2}{k_1}\right)\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_1k_2}+\frac{1}{2}\left(1+\frac{E}{D^2}\right)\left(\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_1k_2}\right)^2. \tag{72}$$

Comparing this relation with the analogous one obtained with a local Eulerian bias expansion (e.g. Fry, Melott & Shandarin 1995; Matarrese, Verde & Heavens 1997), we see that the bispectrum for a set of objects selected by a local Lagrangian bias differ from the results of the local Eulerian bias by the extra inertia term

$$\frac{b_1-1}{2}\left(\frac{k_1}{k_2}+\frac{k_2}{k_1}\right)\frac{\boldsymbol{k}_1\cdot\boldsymbol{k}_2}{k_1k_2}, \tag{73}$$

which implies a different shape dependence.

The halo bispectrum amplitude $Q(\theta)$, at $z=z_{\rm f}=0$, for configurations with sides $k_1=1$, $k_2=1/2$, separated by an angle θ, is shown in Fig. 4, for scale-free models with $n=-2$ and -1, with $\Omega=1$. Two different cases are considered: our local Lagrangian bias model, with linear Eulerian parameters $b_1=2$ and $b_2=1$, and the local Eulerian bias model of Fry & Gaztañaga, with the same Eulerian bias parameters.

Similar reasoning would apply to the skewness, for which the local Lagrangian versus Eulerian bias hypothesis implies a change of the scale dependence, through the extra term

$$-\frac{b_1-1}{(b_1)^2}\,\gamma(\mathscr{R}). \tag{74}$$

With adequate modelling of galaxy formation inside dark matter haloes (e.g. Kauffmann et al. 1997, and references therein) the results of this section can be used to predict the clustering properties of galaxies at different redshifts. In particular, the specific shape dependence of the bispectrum (and related scale dependence of the skewness), implied by our local Lagrangian bias prescription, would reflect into a detectable signature in the statistical properties of the galaxy distribution. Our model therefore provides a valid alternative to local Eulerian bias schemes (e.g. Cen & Ostriker 1992; Coles 1993; Fry & Gaztañaga 1993; Catelan et al. 1994; Mann et al. 1998).

4 CONCLUSIONS

In this paper we studied the non-linear evolution of the clustering of dark matter haloes, using a stochastic approach to single out the halo formation sites directly in Lagrangian space. Our model is based on a local version of the Press–Schechter theory,

which becomes free of the cloud-in-cloud problem after a suitable coarse-graining procedure is applied. The non-linear evolution of the halo distribution is then followed exactly by relating it to the dynamics of the Lagrangian patch of fluid which the nascent halo belongs to.

This formalism allowed us to obtain the bias random field relating the local halo density contrast to the underlying mass distribution. The expression for the halo bias field, reported in equations (48) and (49), represents the most relevant result of our paper. Because of the locality in Lagrangian space inherent in our approach, such a bias field turns out to be non-local in Eulerian coordinates, which has relevant implications for the clustering properties of luminous objects like galaxies and galaxy clusters that formed inside dark matter haloes.

Our method contains two Lagrangian smoothing scales. The scale R, selecting the halo mass, and the background scale $R_0 \gg R$ allow us to define the Lagrangian halo counting field as the local PS mass function in a patch with comoving background density $\rho_b[1+\epsilon_{bg}(\boldsymbol{q}, z)]$, ϵ_{bg} being the linear mass fluctuation smoothed on the scale R_0. Given the role of the latter, it would appear that our description of halo clustering makes sense only on scales larger than R_0. On the other hand, the derivation of the Lagrangian correlation function in Section 2.4, which does not make use of the background field, suggests that we can actually extrapolate our Lagrangian results down to separation comparable to the halo size. This result is further confirmed by an analysis in terms of space-correlated Langevin equations (Porciani et al. 1998). The numerical results of MW and Mo et al. (1996) support the idea that such an extrapolation would apply even in the non-linearly evolved case. In our treatment of the non-linear regime, the background scale R_0 appears with a complementary role. It is the minimum scale ensuring that the nascent haloes are indeed comoving with the Lagrangian fluid patch to which they belong. This would reasonably require that the Lagrangian fluid elements evolve with negligible orbit crossing (e.g. Kofman et al. 1994).

Once again, let us stress that our approach makes no assumptions about the merger rates of the considered objects. The clear distinction between observed and formation redshift, z and z_f, in our approach implies that the instantaneous merging hypothesis, implicit in the standard PS model, as well as any other realistic approximation can be easily accommodated into our scheme as just the way to relate z_f and z.

Our method for evolving the spatial distribution of the haloes is indeed much more general than the specific application we have considered so far. Given any Lagrangian population of objects specified by some set of physical properties $\mathcal{M}$ (like mass and formation threshold in our halo model), with conserved mean comoving number density $\bar{n}_{obj}(\mathcal{M})$ and local Lagrangian density contrast $\delta_{obj}(\boldsymbol{q}|\mathcal{M})$, our results imply that, at any redshift z, their comoving local density in Eulerian space is given by

$$n_{obj}(\boldsymbol{x}, z|\mathcal{M}) = \bar{n}_{obj}(\mathcal{M}) \int d\boldsymbol{q}\,[1+\delta_{obj}(\boldsymbol{q}|\mathcal{M})]\,\delta_D[\boldsymbol{x}-\boldsymbol{x}(\boldsymbol{q}, z)], \tag{75}$$

where $\boldsymbol{x}(\boldsymbol{q}, z) = \boldsymbol{q} + \boldsymbol{S}(\boldsymbol{q}, z)$, and $\boldsymbol{S}(\boldsymbol{q}, z)$ is the displacement vector of the $\boldsymbol{q}$th Lagrangian element. Smoothing the initial gravitational potential on some scale R_0 is again required, so that the objects assigned to the $\boldsymbol{q}$th patch can be sensibly assumed

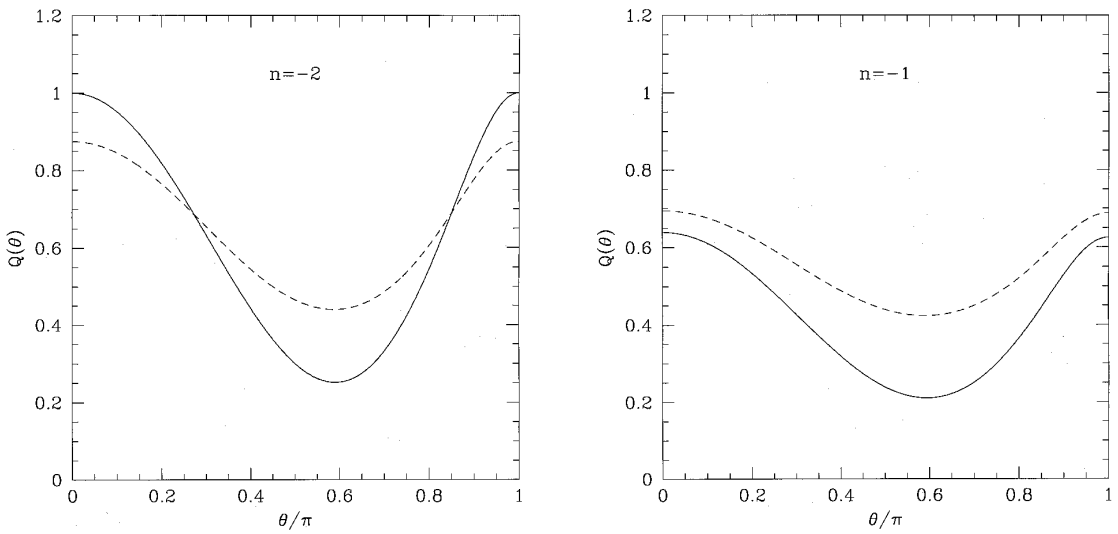

Figure 4. The halo bispectrum amplitude $Q(\theta)$ for configurations with sides $k_1=1$, $k_2=1/2$, separated by an angle θ is plotted versus θ for scale-free models with $n=-2$ and -1 at $z=z_f=0$ and in a flat universe. Two cases are shown for each panel: the local Lagrangian bias model, with linear Eulerian parameters $b_1=2$ and $b_2=1$ (solid line), and the local Eulerian bias model, with the same bias parameters (dashed line).

to be comoving with it. This method could be used, for instance, to follow the clustering of the Lagrangian density maxima in the non-linearly evolved mass density field. This suggests that, starting from low-resolution numerical simulations, one can generate mock catalogues of the given class of objects, with local density correctly specified up to some resolution scale. One can understand the last relation as a local version of the Chapman–Kolmogorov equation of stochastic processes (e.g. van Kampen 1992), stating that the local Eulerian object distribution is the convolution of the Lagrangian object density with the 'conditional particle density', $\delta_D[\boldsymbol{x}-\boldsymbol{x}(\boldsymbol{q}, z)]$, i.e., the probability that a particle is found in $\boldsymbol{x}$ at redshift z, given that it was in $\boldsymbol{q}$ as $z\to\infty$, the only underlying hypothesis being, once again, that these objects move exclusively by the action of gravity. It may be worth noting that the latter equation is actually more general than equation (38), as it also holds in the presence of multistreaming.

Our non-linear stochastic approach can be already considered successful in that, besides recovering the PS mass function, it provides a self-consistent derivative of the Eulerian halo bias, which, to a first approximation, reduces to the MW formula. We, however, also predict both quantitative and qualitative corrections to the MW results, that clearly need to be checked against the outputs of numerical simulations. A definite prediction of our analysis is, for instance, the form of the skewness and of the bispectrum of the spatial halo distribution, which significantly deviates from that deduced with any local Eulerian bias prescription.

ACKNOWLEDGMENTS

The authors greatly acknowledge Alan Heavens, Lev Kofman, Cedric Lacey and Sergei Shandarin for many useful discussions. PC is very grateful to Eric Hivon. PC has been supported by the Danish National Research Foundation at Copenhagen Theoretical Astrophysics Center (TAC), and by the EEC HCMP CT930328 contract at Oxford Astrophysics Department, where part of this investigation was performed. FL, SM and CP thank the Italian MURST for partial financial support. CP is grateful to TAC for kind hospitality.

REFERENCES

Bagla J. S., Padmanabhan T., 1994, MNRAS, 266, 227
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Bernardeau F., 1994, ApJ, 433, 1
Bond J. R., 1988, in Rubin V., Coyne G., eds, Proc. Vatican Study Week, Large-Scale Motions in the Universe. Princeton Univ. Press, Princeton. Vatican City Press, Vatican City
Bond J. R., Myers S. T., 1996, ApJS, 103, 1
Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Bouchet F. R., Juszkiewicz R., Colombi S., Pellat R., 1992, ApJ, 394, L5
Bower R. G., 1991, MNRAS, 248, 332
Brainerd T. G., Scherrer R. J., Villumsen J. V., 1993, ApJ, 418, 570
Catelan P., Coles P., Matarrese S., Moscardini L., 1994, MNRAS, 268, 966
Catelan P., Lucchin F., Matarrese S., Moscardini L., 1995, MNRAS, 276, 39
Cavaliere A., Colafrancesco S., Menci N., 1993, ApJ, 415, 50
Cen R. Y., Ostriker J. P., 1992, ApJ, 399, L113
Colafrancesco S., Lucchin F., Matarrese S., 1989, ApJ, 345, 3
Cole S., 1991, ApJ, 367, 45
Cole S., Kaiser N., 1989, MNRAS, 237, 1127
Coles P., 1993, MNRAS, 262, 1065
Coles P., Jones B., 1991, MNRAS, 248, 1
Coles P., Melott A. L., Shandarin S. F., 1993, MNRAS, 260, 765
Dekel A., 1986, Comments Astrophys., 11, 235
Dekel A., Rees M. J., 1987, Nat, 326, 455
Doroshkevich A. G., 1967, Astrofizika, 3, 175 (Astrophysics, 3, 84)
Doroshkevich A. G., 1970, Astrofizika, 6, 581 (Astrophysics, 6, 320)
Efstathiou G., Frenk C. S., White S. D. M., Davis M., 1988, MNRAS, 235, 715
Fry J. N., 1984, ApJ, 279, 499
Fry J. N., 1996, ApJ, 461, L65
Fry J. N., Gaztañaga E., 1993, ApJ, 413, 447
Fry J. N., Melott A. L., Shandarin S. F., 1995, MNRAS, 274, 745
Gelb J. M., Bertschinger E., 1994, ApJ, 436, 467
Hui L., Bertschinger E., 1996, ApJ, 471, 1
Kaiser N., 1984, ApJ, 284, L9
Kashlinsky A., 1987, ApJ, 317, 19
Kauffmann G., Nusser A., Steinmetz M., 1997, MNRAS, 286, 795
Kofman L., 1991, in Sato K., ed., Primordial Nucleosynthesis and Evolution of the Early Universe. Kluwer Academic Press, Dordrecht
Kofman L., Pogosyan D., Shandarin S. F., Melott A. L., 1992, ApJ, 393, 437
Kofman L., Bertschinger E., Gelb J. M., Nusser A., Dekel A., 1994, ApJ, 420, 44
Lacey C., Cole S., 1993, MNRAS, 262, 627
Lacey C., Cole S., 1994, MNRAS, 271, 676
Lee J., Shandarin S. F., 1997, Bull. AAS, 191, 8707
Mann R. G., Peacock J. A., Heavens A. F., 1998, MNRAS, 293, 209
Manrique A., Salvador-Solé E., 1995, ApJ, 453, 6
Manrique A., Salvador-Solé E., 1996, ApJ, 467, 504
Matarrese S., Lucchin F., Moscardini L., Saez D., 1992, MNRAS, 259, 437
Matarrese S., Coles P., Lucchin F., Moscardini L., 1997, MNRAS, 286, 115
Matarrese S., Verde L., Heavens A. F., 1997, MNRAS, 290, 651
Mo H. J., White S. D. M., 1996, MNRAS, 282, 347 (MW)
Mo H. J., Jing Y. P., White S. D. M., 1996, MNRAS, 282, 1096
Monaco P., 1995, ApJ, 447, 23
Nusser A., Davis M., 1994, ApJ, 421, L1
Peacock J. A., 1997, MNRAS, 284, 885
Peacock J. A., Dodds S. J., 1994, MNRAS, 267, 1020
Peacock J. A., Heavens A. F., 1990, MNRAS, 243, 133

Porciani C., Matarrese S., Lucchin F., Catelan P., 1998, MNRAS, in press
Press W. H., Schechter P., 1974, ApJ, 187, 425 (PS)
Sathyaprakash B. S., Sahni V., Munshi D., Pogosyan D., Melott A. L., 1995, MNRAS, 275, 463
Shandarin S. F., 1980, Astrofizika, 16, 769 (Astrophysics, 16, 439)
Shandarin S. F., Zel'dovich Ya. B., 1984, Phys. Rev. Lett., 52, 1488
van Kampen N. G., 1992, Stochastic Processes in Physics and Chemistry, North-Holland, Amsterdam
Zel'dovich Ya. B., 1970, A&A, 5, 84

APPENDIX A: GROWTH FACTORS IN FRIEDMANN UNIVERSE MODELS

The expressions for the first- and second-order growth factors $D(z)$ and $E(z)$ have not been given in the main text. An easy derivation can be given following Shandarin (1980) and using the relation

$$\Omega^{-1}-1=(\Omega_0^{-1}-1)(1+z)^{-1}. \tag{A1}$$

We consider only cases with vanishing cosmological constant. The growth factor $\mathscr{D}(z;\Omega_0)$ of linear density perturbations reads, for the different geometries,

$$\mathscr{D}(z;\Omega_0)=\begin{cases}\dfrac{5}{2}+\dfrac{15}{2}\,\dfrac{\Omega_0(1+z)}{1-\Omega_0}\left[1-\dfrac{1}{2}\sqrt{\dfrac{1+\Omega_0 z}{1-\Omega_0}}\ln\left(\dfrac{\Omega_0(1+z)}{2-\Omega_0(1-z)-2\sqrt{(1-\Omega_0)(1+\Omega_0 z)}}\right)\right] & (\Omega_0<1)\\ (1+z)^{-1} & (\Omega_0=1)\\ -\dfrac{5}{2}+\dfrac{15}{2}\,\dfrac{\Omega_0(1+z)}{\Omega_0-1}\left[1+\sqrt{\dfrac{1+\Omega_0 z}{\Omega_0-1}}\arctan\left(-\sqrt{\dfrac{\Omega_0-1}{1+\Omega_0 z}}\right)\right] & (\Omega_0>1).\end{cases} \tag{A2}$$

The expressions for the second-order growth factors $\mathscr{E}(z;\Omega_0)$ are slightly more cumbersome:

$$\mathscr{E}(z;\Omega_0)=-\frac{25}{8}-\frac{225}{8}\,\frac{\Omega_0(1+z)}{1-\Omega_0}\left[\left[1-\frac{1}{2}\sqrt{\frac{1+\Omega_0 z}{1-\Omega_0}}\ln\left[\frac{\Omega_0(1+z)}{2-\Omega_0(1-z)-2\sqrt{(1-\Omega_0)(1+\Omega_0 z)}}\right]\right.\right.$$
$$\left.\left.+\frac{1}{2}\left\{-\sqrt{\frac{1+\Omega_0 z}{1-\Omega_0}}+\frac{1}{2}\,\frac{\Omega_0(1+z)}{1-\Omega_0}\ln\left[\frac{\Omega_0(1+z)}{2-\Omega_0(1-z)-2\sqrt{(1-\Omega_0)(1+\Omega_0 z)}}\right]\right\}^2\right]\right] \qquad (\Omega_0<1), \tag{A3}$$

$$\mathscr{E}(z;\Omega_0)=-\frac{3}{7(1+z)^2} \qquad (\Omega_0=1), \tag{A4}$$

$$\mathscr{E}(z;\Omega_0)=-\frac{25}{8}+\frac{225}{8}\,\frac{\Omega_0(1+z)}{\Omega_0-1}\left\{1+\sqrt{\frac{1+\Omega_0 z}{\Omega_0-1}}\arctan\left(-\sqrt{\frac{\Omega_0-1}{1+\Omega_0 z}}\right)\right.$$
$$\left.+\frac{1}{2}\left[\sqrt{\frac{1+\Omega_0 z}{\Omega_0-1}}+\frac{\Omega_0(1+z)}{\Omega_0-1}\arctan\left(-\sqrt{\frac{\Omega_0-1}{1+\Omega_0 z}}\right)\right]^2\right\} \qquad (\Omega_0>1). \tag{A5}$$

Notice that we are implicitly adopting here the normalization suggested by Shandarin (1980), so that, in the limit $z\to\infty$ one recovers the Einstein–de Sitter case, $\mathscr{D}(z;\Omega_0)\to(1+z)^{-1}$. However, in the main text we normalized to unity the linear growing factors extrapolated to the present time; so, for any geometry, we define $D(z)\equiv\mathscr{D}(z;\Omega_0)/\mathscr{D}(z=0;\Omega_0)$ and $E(z)\equiv\mathscr{E}(z;\Omega_0)/[\mathscr{D}(z=0;\Omega_0)]^2$.